

The Unbounded Benefit of Encoder Cooperation for the k -user MAC

Parham Noorzad, *Student Member, IEEE*, Michelle Effros, *Fellow, IEEE*,
and Michael Langberg, *Senior Member, IEEE*

Abstract

Cooperation strategies allow communication devices to work together to improve network capacity. Consider a network consisting of a k -user multiple access channel (MAC) and a node that is connected to all k encoders via rate-limited bidirectional links, referred to as the “cooperation facilitator” (CF). Define the cooperation benefit as the sum-capacity gain resulting from the communication between the encoders and the CF and the cooperation rate as the total rate the CF shares with the encoders. This work demonstrates the existence of a class of k -user MACs where the ratio of the cooperation benefit to cooperation rate tends to infinity as the cooperation rate tends to zero. Examples of channels in this class include the binary erasure MAC for $k = 2$ and the k -user Gaussian MAC for any $k \geq 2$.

Index Terms

Conferencing encoders, cooperation facilitator, cost constraints, edge removal problem, multiple access channel, multivariate covering lemma, network information theory.

I. INTRODUCTION

In large networks, resources may not always be distributed evenly across the network. There may be times where parts of a network are underutilized, while others are overconstrained, leading to suboptimal performance. In such situations, end users are not able to use their devices to their full capabilities.

One approach to address this problem allows some nodes in the network to “cooperate,” that is, work together, either directly or indirectly, to achieve common goals. The model we next introduce is based on this idea.

In the classical k -user multiple access channel (MAC) [3], there are k encoders and a single decoder. Each encoder has a private message which it transmits over n channel uses to the decoder. The decoder, once it receives n output symbols, finds the messages of all k encoders with small average probability of error. In this model, the encoders cannot cooperate, since each encoder only has access to its own message.

This paper was presented in part at the 2015 IEEE International Symposium of Information Theory (ISIT) in Hong Kong [1] and the 2016 IEEE ISIT in Barcelona, Spain [2].

This material is based upon work supported by the National Science Foundation under Grant Numbers 15727524, 1526771, and 1321129.

P. Noorzad and M. Effros are with the California Institute of Technology, Pasadena, CA 91125 USA (emails: parham@caltech.edu, effros@caltech.edu).

M. Langberg is with the State University of New York at Buffalo, Buffalo, NY 14260 USA (email: mikel@buffalo.edu).

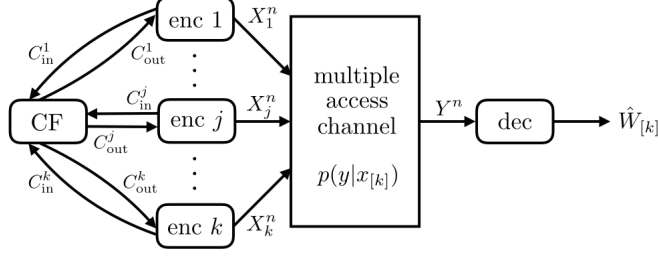


Figure 1. The network consisting of a k -user MAC and a CF. For $j \in [k]$, encoder j has access to message $w_j \in [2^{nR_j}]$.

We now consider an alternative scenario where our k -user MAC is part of a larger network. In this network, there is a node that is connected to all k encoders and acts as a “cooperation facilitator” (CF). Specifically, for every $j \in [k]$,¹ there is a link of capacity $C_{\text{in}}^j \geq 0$ going from encoder j to the CF and a link of capacity $C_{\text{out}}^j \geq 0$ going back. The CF helps the encoders exchange information before they transmit their codewords over the MAC. Figure 1 depicts a network consisting of a k -user MAC and a $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$ -CF, where $\mathbf{C}_{\text{in}} = (C_{\text{in}}^j)_{j \in [k]}$ and $\mathbf{C}_{\text{out}} = (C_{\text{out}}^j)_{j \in [k]}$ denote the capacities of the CF input and output links. In this figure, $X_{[k]}^n = (X_1^n, \dots, X_k^n)$ is the vector of the channel inputs of the k encoders, and $\hat{W}_{[k]} = (\hat{W}_1, \dots, \hat{W}_k)$ is the vector of message reproductions at the decoder.

The communication between the CF and the encoders occurs over a number of rounds. In the first round of cooperation, each encoder sends a rate-limited function of its message to the CF, and the CF sends a rate-limited function of what it receives back to each encoder. Communication between the encoders and the CF may continue for a finite number of rounds, with each node potentially using information received in prior rounds to determine its next transmission. Once the communication between the CF and the encoders is done, each encoder uses its message and what it has learned through the CF to choose a codeword, which it transmits across the channel.

Our main result (Theorem 3) determines a set of MACs where the benefit of encoder cooperation through a CF grows very quickly with \mathbf{C}_{out} . Specifically, we find a class of MACs \mathcal{C}^* , where every MAC in \mathcal{C}^* has the property that for any fixed $\mathbf{C}_{\text{in}} \in \mathbb{R}_{>0}^k$, the sum-capacity of that MAC with a $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$ -CF has an infinite derivative in the direction of every $\mathbf{v} \in \mathbb{R}_{>0}^k$ at $\mathbf{C}_{\text{out}} = \mathbf{0}$. In other words, as a function of \mathbf{C}_{out} , the sum-capacity grows faster than any function with bounded derivative at $\mathbf{C}_{\text{out}} = \mathbf{0}$. This means that for any MAC in \mathcal{C}^* , sharing a small number of bits with each encoder leads to a large gain in sum-capacity.

An important implication of this result is the existence of a memoryless network that does not satisfy the “edge removal property” [4], [5]. A network satisfies the edge removal property if removing an edge of capacity $\delta > 0$ changes the capacity region by at most δ in each dimension. Thus removing an edge of capacity δ from a network which has k sources and satisfies the edge removal property, decreases sum-capacity by at most $k\delta$, a linear function of δ . Now consider a network consisting of a MAC in \mathcal{C}^* and a $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$ -CF, where $\mathbf{C}_{\text{in}} \in \mathbb{R}_{>0}^k$. Our main result (Theorem 3) implies that for small \mathbf{C}_{out} , removing all the output edges reduces sum-capacity by an amount much

¹The notation $[x]$ describes the set $\{1, \dots, \lfloor x \rfloor\}$ for any real number $x \geq 1$.

larger than $k \sum_{j \in [k]} C_{\text{out}}^j$. Thus there exist memoryless networks that do not satisfy the edge removal property. The first example of such a network appeared in [6].

We introduce the coding scheme that leads to Theorem 3 in Section IV. This scheme combines forwarding, coordination, and classical MAC coding. In forwarding, each encoder sends part of its message to all other encoders by passing that information through the CF.² When $k = 2$, forwarding is equivalent to a single round of conferencing as described in [8]. The coordination strategy is a modified version of Marton's coding scheme for the broadcast channel [9], [10]. To implement this strategy, the CF shares information with the encoders that enables them to transmit codewords that are jointly typical with respect to a *dependent* distribution; this is proven using a multivariate version of the covering lemma [11, p. 218]. The multivariate covering lemma is stated for strongly typical sets in [11]. In Appendix A, using the proof of the 2-user case from [11] and techniques from [12], we prove this lemma for weakly typical sets [13, p. 251]. Using weakly typical sets in our achievability proof allows our results to extend to continuous (e.g., Gaussian) channels without the need for quantization. Finally, the classical MAC strategy is Ulrey's [3] extension of Ahlswede's [14], [15] and Liao's [16] coding strategy to the k -user MAC.

Using techniques from Willems [8], we derive an outer bound (Proposition 5) for the capacity region of the MAC with a $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$ -CF. This outer bound does not capture the dependence of the capacity region on \mathbf{C}_{out} and is thus loose for some values of \mathbf{C}_{out} . However, if the entries of \mathbf{C}_{out} are sufficiently larger than the entries of \mathbf{C}_{in} , then our inner and outer bounds agree and we obtain the capacity region (Corollary 6).

In Section V, we apply our results to the 2-user Gaussian MAC with a CF that has access to the messages of both encoders and has links of output capacity C_{out} . We show that for small C_{out} , the achievable sum-rate approximately equals a constant times $\sqrt{C_{\text{out}}}$. A similar approximation holds for a weighted version of the sum-rate as well, as we see in Proposition 7. This result implies that at least for the 2-user Gaussian MAC, the benefit of cooperation is not limited to sum-capacity and applies to other capacity region metrics as well.

In Section VI, we consider the extension of Willems' conferencing model [8] from 2 to k users. A special case of this model with $k = 3$ is studied in [17] for the Gaussian MAC. While the authors of [17] use two conferencing rounds in their achievability result, it is not clear from [17] if there is a benefit in using two rounds instead of one, and if so, how large that benefit is. Here we explicitly show that a single conferencing round is not optimal for $k \geq 3$, even though it is known to be optimal when $k = 2$ [8]. Finally, we apply our outer bound for the k -user MAC with a CF to obtain an outer bound for the k -user MAC with conferencing. The resulting outer bound is tight when $k = 2$.

In the next section, we formally define the capacity region of the network consisting of a k -user MAC and a CF.

II. MODEL

Consider a network with k encoders, a CF, a k -user MAC, and a decoder (Figure 1). For each $j \in [k]$, encoder j communicates with the CF using noiseless links of capacities $C_{\text{in}}^j \geq 0$ and $C_{\text{out}}^j \geq 0$ going to and from the CF,

²While it is possible to consider encoders that send *different* parts of their messages to different encoders using Han's result for the MAC with correlated sources [7], we avoid these cases for simplicity.

respectively. The k encoders communicate with the decoder through a MAC $(\mathcal{X}_{[k]}, p(y|x_{[k]}), \mathcal{Y})$, where

$$\mathcal{X}_{[k]} = \prod_{j=1}^k \mathcal{X}_j,$$

and an element of $\mathcal{X}_{[k]}$ is denoted by $x_{[k]}$. We say a MAC is discrete if $\mathcal{X}_{[k]}$ and \mathcal{Y} are either finite or countably infinite, and $p(y|x_{[k]})$ is a probability mass function on \mathcal{Y} for every $x_{[k]} \in \mathcal{X}_{[k]}$. We say a MAC is continuous if $\mathcal{X}_{[k]} = \mathbb{R}^k$, $\mathcal{Y} = \mathbb{R}$, and $p(y|x_{[k]})$ is a probability density function on \mathcal{Y} for all $x_{[k]}$. In addition, we assume that our channel is memoryless and without feedback [13, p. 193], so that for every positive integer n , the n th extension channel of our MAC is given by $p(y^n|x_{[k]}^n)$, where

$$\forall (x_{[k]}^n, y^n) \in \mathcal{X}_{[k]}^n \times \mathcal{Y}^n : p(y^n|x_{[k]}^n) = \prod_{t=1}^n p(y_t|x_{[k]}^n_t).$$

An example of a continuous MAC is the k -user Gaussian MAC with noise variance $N > 0$, where

$$p(y|x_{[k]}) = \frac{1}{\sqrt{2\pi N}} \exp \left[-\frac{1}{2N} \left(y - \sum_{j \in [k]} x_j \right)^2 \right] \quad (1)$$

Henceforth, all MACs are memoryless and without feedback, and either discrete or continuous.

We next describe a

$$((2^{nR_1}, \dots, 2^{nR_k}), n, L)\text{-code}$$

for the MAC $(\mathcal{X}_{[k]}, p(y|x_{[k]}), \mathcal{Y})$ with a $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$ -CF with cost functions $(b_j)_{j \in [k]}$ and cost constraint vector $\mathbf{B} = (B_j)_{j \in [k]} \in \mathbb{R}_{\geq 0}^k$. For each $j \in [k]$, cost function b_j is a fixed mapping from \mathcal{X}_j to $\mathbb{R}_{\geq 0}$. Each encoder $j \in [k]$ wishes to transmit a message $w_j \in [2^{nR_j}]$ to the decoder. This is accomplished by first exchanging information with the CF and then transmitting across the MAC. Communication with the CF occurs in L rounds. For each $j \in [k]$ and $\ell \in [L]$, sets $\mathcal{U}_{j\ell}$ and $\mathcal{V}_{j\ell}$, respectively, describe the alphabets of symbols that encoder j can send to and receive from the CF in round ℓ . These alphabets satisfy the link capacity constraints

$$\begin{aligned} \sum_{\ell=1}^L \log |\mathcal{U}_{j\ell}| &\leq nC_{\text{in}}^j \\ \sum_{\ell=1}^L \log |\mathcal{V}_{j\ell}| &\leq nC_{\text{out}}^j. \end{aligned} \quad (2)$$

The operation of encoder j and the CF, respectively, in round ℓ are given by

$$\begin{aligned} \varphi_{j\ell} : [2^{nR_j}] \times \mathcal{V}_j^{\ell-1} &\rightarrow \mathcal{U}_{j\ell} \\ \psi_{j\ell} : \prod_{i=1}^k \mathcal{U}_i^\ell &\rightarrow \mathcal{V}_{j\ell}. \end{aligned}$$

where $\mathcal{U}_j^\ell = \prod_{\ell'=1}^\ell \mathcal{U}_{j\ell'}$ and $\mathcal{V}_j^\ell = \prod_{\ell'=1}^\ell \mathcal{V}_{j\ell'}$. After its exchange with the CF, encoder j applies a function

$$f_j : [2^{nR_j}] \times \mathcal{V}_j^L \rightarrow \mathcal{X}_j^n,$$

to choose a codeword, which it transmits across the channel. In addition, every x_j^n in the range of f_j satisfies

$$\sum_{t=1}^n b_j(x_{jt}) \leq nB_j.$$

The decoder receives channel output Y^n and applies

$$g : \mathcal{Y}^n \rightarrow \prod_{j=1}^k [2^{nR_j}]$$

to obtain estimate $\hat{W}_{[k]}$ of the message vector $w_{[k]}$.

The encoders, CF, and decoder together define a

$$((2^{nR_1}, \dots, 2^{nR_k}), n, L)\text{-code}.$$

The average error probability of the code is $P_e^{(n)} = \Pr \{g(Y^n) \neq W_{[k]}\}$, where $W_{[k]}$ is the transmitted message vector and is uniformly distributed on $\prod_{j=1}^k [2^{nR_j}]$. A rate vector $R_{[k]} = (R_1, \dots, R_k)$ is *achievable* if there exists a sequence of $((2^{nR_1}, \dots, 2^{nR_k}), n, L)$ codes with $P_e^{(n)} \rightarrow 0$ as $n \rightarrow \infty$. The capacity region, $\mathcal{C}(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$, is defined as the closure of the set of all achievable rate vectors.

III. RESULTS

In this section, we describe the key results. In Subsection III-A, we present our inner bound. In Subsection III-B, we state our main result, which proves the existence of a class of MACs with large cooperation gain. Finally, in Subsection III-C, we discuss our outer bound.

A. Inner Bound

Using the coding scheme we introduce in Section IV, we obtain an inner bound for the capacity region of the k -user MAC with a $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$ -CF. The following definitions are useful for describing that bound. Choose vectors $\mathbf{C}_0 = (C_{j0})_{j=1}^k$ and $\mathbf{C}_d = (C_{jd})_{j=1}^k$ in $\mathbb{R}_{\geq 0}^k$ such that for all $j \in [k]$,

$$C_{j0} \leq C_{\text{in}}^j \tag{3}$$

$$C_{jd} + \sum_{i \neq j} C_{i0} \leq C_{\text{out}}^j. \tag{4}$$

Here C_{j0} is the number of bits per channel use encoder j sends directly to the other encoders via the CF and C_{jd} is the number of bits per channel use the CF transmits to encoder j to implement the coordination strategy. Subscript “ d ” in C_{jd} alludes to the dependence created through coordination. Let $S_d = \{j \in [k] : C_{jd} \neq 0\}$ be the set of encoders that participate in this dependence.

Fix alphabets $\mathcal{U}_0, \mathcal{U}_1, \dots, \mathcal{U}_k$. For every nonempty $S \subseteq [k]$, let \mathcal{U}_S be the set of all $u_S = (u_j)_{j \in S}$ where $u_j \in \mathcal{U}_j$ for all $j \in S$. Define the set \mathcal{X}_S similarly. Let $\mathcal{P}(\mathcal{U}_0, \mathcal{U}_{[k]}, \mathcal{X}_{[k]}, S_d)$ be the set of all distributions on $\mathcal{U}_0 \times \mathcal{U}_{[k]} \times \mathcal{X}_{[k]}$ that are of the form

$$p(u_0) \cdot \prod_{i \in S_d^c} p(u_i | u_0) \cdot p(u_{S_d} | u_0, u_{S_d^c}) \cdot \prod_{j \in [k]} p(x_j | u_0, u_j), \tag{5}$$

satisfy the dependence constraints³

$$\zeta_S := \sum_{j \in S} C_{jd} - \sum_{j \in S} H(U_j | U_0) + H(U_S | U_0, U_{S_d^c}) > 0 \quad \forall \emptyset \subsetneq S \subseteq S_d,$$

³The constraint on ζ_S is imposed by the multivariate covering lemma (Appendix A), which we use in the proof of our inner bound.

and cost constraints

$$\mathbb{E}[b_j(X_j)] \leq B_j \quad \forall j \in [k]. \quad (6)$$

Here U_0 encodes the “common message,” which, for every $j \in [k]$, contains nC_{j0} bits from the message of encoder j and is shared with all other encoders through the CF; each random variable U_j captures the information encoder j receives from the CF to create dependence with the codewords of other encoders. The random variable X_j represents the symbol encoder j transmits over the channel.

For any $\mathbf{C}_0, \mathbf{C}_d \in \mathbb{R}_{\geq 0}^k$ satisfying (3) and (4) and any $p \in \mathcal{P}(\mathcal{U}_0, \mathcal{U}_{[k]}, \mathcal{X}_{[k]}, S_d)$, let $\mathcal{R}(\mathbf{C}_0, \mathbf{C}_d, p)$ be the set of all (R_1, \dots, R_k) for which

$$\sum_{j \in [k]} R_j < I(X_{[k]}; Y) - \zeta_{S_d}, \quad (7)$$

and for every $S, T \subseteq [k]$,

$$\begin{aligned} \sum_{j \in A} (R_j - C_{j0})^+ + \sum_{j \in B \cap T} (R_j - C_{\text{in}}^j)^+ \\ < I(U_A, X_{A \cup (B \cap T)}; Y | U_0, U_B, X_{B \setminus T}) - \zeta_{(A \cup B) \cap S_d} \end{aligned} \quad (8)$$

holds for some sets A and B for which $S \cap S_d^c \subseteq A \subseteq S$ and $S^c \cap S_d^c \subseteq B \subseteq S^c$.

We next state our inner bound for the k -user MAC with encoder cooperation via a CF. The coding strategy that achieves this inner bound uses only a single round of cooperation ($L = 1$). The proof is given in Subsection VII-A.

Theorem 1 (Inner Bound). *For any MAC $(\mathcal{X}_{[k]}, p(y|x_{[k]}), \mathcal{Y})$ with a $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$ -CF,*

$$\mathcal{C}(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}}) \supseteq \overline{\bigcup \mathcal{R}(\mathbf{C}_0, \mathbf{C}_d, p)}$$

where \bar{A} denotes the closure of set A and the union is over all \mathbf{C}_0 and \mathbf{C}_d satisfying (3) and (4), and $p \in \mathcal{P}(\mathcal{U}_0, \mathcal{U}_{[k]}, \mathcal{X}_{[k]}, S_d)$.

The achievable region given in Theorem 1 is convex and thus we do not require the convex hull operation. The proof is similar to [1], [18] and is omitted.

The next corollary treats the case where the CF transmits the bits it receives from each encoder to all other encoders without change. In this case, our coding strategy simply combines forwarding with classical MAC encoding. We obtain this result from Theorem 1 by setting $C_{jd} = 0$ and $|\mathcal{U}_j| = 1$ for all $j \in [k]$ and choosing $A = S$ and $B = S^c$ for every $S, T \subseteq [k]$. In Corollary 2, $\mathcal{P}_{\text{ind}}(\mathcal{U}_0, \mathcal{X}_{[k]})$ is the set of all distributions $p(u_0) \prod_{j \in [k]} p(x_j | u_0)$ that satisfy the cost constraints (6).

Corollary 2 (Forwarding Inner Bound). *The capacity region of any MAC with a $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$ -CF contains the set of all rate vectors that for some constants $(C_{j0})_{j \in [k]}$ (satisfying (3) and (4) with $C_{jd} = 0$ for all j) and some distribution $p \in \mathcal{P}_{\text{ind}}(\mathcal{U}_0, \mathcal{X}_{[k]})$, satisfy*

$$\begin{aligned} \sum_{j \in S} R_j &< I(X_S; Y | U_0, X_{S^c}) + \sum_{j \in S} C_{j0} \quad \forall \emptyset \neq S \subseteq [k] \\ \sum_{j \in [k]} R_j &< I(X_{[k]}; Y). \end{aligned}$$

B. Sum-Capacity Gain

We wish to understand when cooperation leads to a benefit that exceeds the resources employed to enable it. Therefore, we compare the gain in sum-capacity obtained through cooperation to the number of bits shared with the encoders to enable that gain.

For any k -user MAC with a $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$ -CF, define the sum-capacity as

$$C_{\text{sum}}(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}}) = \max_{\mathcal{C}(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})} \sum_{j=1}^k R_j.$$

For a fixed $\mathbf{C}_{\text{in}} \in \mathbb{R}_{\geq 0}^k$, define the “sum-capacity gain” $G : \mathbb{R}_{\geq 0}^k \rightarrow \mathbb{R}_{\geq 0}$ as

$$G(\mathbf{C}_{\text{out}}) = C_{\text{sum}}(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}}) - C_{\text{sum}}(\mathbf{C}_{\text{in}}, \mathbf{0}),$$

where $\mathbf{C}_{\text{out}} = (C_{\text{out}}^j)_{j=1}^k$ and $\mathbf{0} = (0, \dots, 0)$. Note that regardless of \mathbf{C}_{in} , it follows from (2) that no cooperation is possible when $\mathbf{C}_{\text{out}} = \mathbf{0}$. Thus

$$C_{\text{sum}}(\mathbf{C}_{\text{in}}, \mathbf{0}) = C_{\text{sum}}(\mathbf{0}, \mathbf{0}) = \max_{p \in \mathcal{P}_{\text{ind}}(X_{[k]})} I(X_{[k]}; Y),$$

where $\mathcal{P}_{\text{ind}}(X_{[k]})$ is the set of all independent distributions

$$p(x_{[k]}) = \prod_{j \in [k]} p(x_j)$$

on $\mathcal{X}_{[k]}$ that satisfy the cost constraints (6). Similarly, $\mathcal{P}(\mathcal{X}_{[k]})$ is the set of *all* distributions on $\mathcal{X}_{[k]}$ that satisfy (6).

For sets $\mathcal{X}_1, \dots, \mathcal{X}_k, \mathcal{Y}$, cost functions $(b_j)_{j \in [k]}$, and cost constraints $(B_j)_{j \in [k]}$, we next define a special class of MACs $\mathcal{C}^*(\mathcal{X}_{[k]}, \mathcal{Y})$. We say a MAC $(\mathcal{X}_{[k]}, p(y|x_{[k]}), \mathcal{Y})$ is in $\mathcal{C}^*(\mathcal{X}_{[k]}, \mathcal{Y})$, if there exists $p_{\text{ind}} \in \mathcal{P}_{\text{ind}}(X_{[k]})$ that satisfies

$$I_{\text{ind}}(X_{[k]}; Y) = \max_{p \in \mathcal{P}_{\text{ind}}(X_{[k]})} I(X_{[k]}; Y),$$

and $p_{\text{dep}} \in \mathcal{P}(X_{[k]})$ whose support is contained in the support of p_{ind} and satisfies

$$I_{\text{dep}}(X_{[k]}; Y) + D(p_{\text{dep}}(y) \| p_{\text{ind}}(y)) > I_{\text{ind}}(X_{[k]}; Y). \quad (9)$$

In the above equation, $p_{\text{dep}}(y)$ and $p_{\text{ind}}(y)$ are the output distributions corresponding to the input distributions $p_{\text{dep}}(x_{[k]})$ and $p_{\text{ind}}(x_{[k]})$, respectively. We remark that (9) is equivalent to

$$\mathbb{E}_{\text{dep}} \left[D(p(y|X_{[k]}) \| p_{\text{ind}}(y)) \right] > \mathbb{E}_{\text{ind}} \left[D(p(y|X_{[k]}) \| p_{\text{ind}}(y)) \right],$$

where the expectations are with respect to $p_{\text{dep}}(x_{[k]})$ and $p_{\text{ind}}(x_{[k]})$, respectively.

Using these definitions, we state our main result which captures a family of MACs for which the slope of the gain function is infinite in every direction at $\mathbf{C}_{\text{out}} = \mathbf{0}$. In this statement, for any unit vector $\mathbf{v} \in \mathbb{R}_{\geq 0}^k$, $D_{\mathbf{v}}G$ is the directional derivative of G in the direction of \mathbf{v} . The proof appears in Subsection VII-B.

Theorem 3 (Sum-capacity). *Let $(\mathcal{X}_{[k]}, p(y|x_{[k]}), \mathcal{Y})$ be a MAC in $\mathcal{C}^*(\mathcal{X}_{[k]}, \mathcal{Y})$ and $\mathbf{C}_{\text{in}} \in \mathbb{R}_{> 0}^k$. Then for any unit vector $\mathbf{v} \in \mathbb{R}_{> 0}^k$,*

$$(D_{\mathbf{v}}G)(\mathbf{0}) = \infty.$$

Note that for continuous MACs, when for $j \in [k]$ and $x \in \mathbb{R}$, $b_j(x) = x^2$, cost constraints are referred to as power constraints. In addition, for every $j \in [k]$, the variable P_j is commonly used instead of B_j . Our next proposition provides necessary and sufficient conditions under which the k -user Gaussian MAC with power constraints is in $\mathcal{C}^*(\mathbb{R}^k, \mathbb{R})$. The proof is provided in Subsection VII-C.

Proposition 4. *The k -user Gaussian MAC with power constraint vector $\mathbf{P} = (P_j)_{j \in [k]} \in \mathbb{R}_{\geq 0}^k$ is in $\mathcal{C}^*(\mathbb{R}^k, \mathbb{R})$ if and only if at least two entries of \mathbf{P} are positive.*

C. Outer Bound

We next describe our outer bound. While we only make use of a single round of cooperation in our inner bound (Theorem 1), the outer bound applies to all coding schemes regardless of the number of rounds.

Proposition 5 (Outer Bound). *For the MAC $(\mathcal{X}_{[k]}, p(y|x_{[k]}), \mathcal{Y})$, $\mathcal{C}(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$ is a subset of the set of all rate vectors that for some distribution $p \in \mathcal{P}_{\text{ind}}(\mathcal{U}_0, \mathcal{X}_{[k]})$ satisfy*

$$\sum_{j \in S} R_j \leq I(X_S; Y | U_0, X_{S^c}) + \sum_{j \in S} C_{\text{in}}^j \quad \forall \emptyset \neq S \subseteq [k] \quad (10)$$

$$\sum_{j \in [k]} R_j \leq I(X_{[k]}; Y). \quad (11)$$

The proof of this proposition is given in Subsection VII-D. Our proof uses ideas similar to the proof of the converse for the 2-user MAC with conferencing [8].

If the capacities of the CF output links are sufficiently large, our inner and outer bounds coincide and we obtain the capacity region. This follows by setting $C_{j0} = C_{\text{in}}^j$ for all $j \in [k]$ in our forwarding inner bound (Corollary 2) and comparing it with the outer bound given in Proposition 5.

Corollary 6. *For the MAC $(\mathcal{X}_{[k]}, p(y|x_{[k]}), \mathcal{Y})$ with a $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$ -CF, if*

$$\forall j \in [k] : C_{\text{out}}^j \geq \sum_{i: i \neq j} C_{\text{in}}^i,$$

then our inner and outer bounds agree.

IV. THE CODING SCHEME

Choose nonnegative constants $(C_{j0})_{j=1}^k$ and $(C_{jd})_{j=1}^k$ such that (3) and (4) hold for all $j \in [k]$. Fix a distribution $p \in \mathcal{P}(\mathcal{U}_0, \mathcal{U}_{[k]}, \mathcal{X}_{[k]}, S_d)$ and constants $\epsilon, \delta > 0$. Let

$$R_{j0} = \min\{R_j, C_{j0}\}$$

$$R_{jd} = \min\{R_j, C_{\text{in}}^j\} - R_{j0}$$

$$R_{jj} = R_j - R_{j0} - R_{jd} = (R_j - C_{\text{in}}^j)^+,$$

where $x^+ = \max\{x, 0\}$ for any real number x . For every $j \in [k]$, split the message of encoder j as $w_j = (w_{j0}, w_{jd}, w_{jj})$, where $w_{j0} \in [2^{nR_{j0}}]$, $w_{jd} \in [2^{nR_{jd}}]$, $w_{jj} \in [2^{nR_{jj}}]$. For all $j \in [k]$, encoder j sends (w_{j0}, w_{jd})

noiselessly to the CF. This is possible, since $R_{j0} + R_{jd}$ is less than or equal to C_{in}^j . The CF sends w_{j0} to all other encoders via its output links and uses w_{jd} to implement the coordination strategy to be described below. Due to the CF rate constraints, encoder j cannot share the remaining part of its message, w_{jj} , with the CF. Instead, it transmits w_{jj} over the channel using the classical MAC strategy.

Let $\mathcal{W}_0 = \prod_{j=1}^k [2^{nR_{j0}}]$. For every $w_0 \in \mathcal{W}_0$, let $U_0^n(w_0)$ be drawn independently according to

$$\Pr \{U_0^n(w_0) = u_0^n\} = \prod_{t=1}^n p(u_{0t}).$$

Given $U_0^n(w_0) = u_0^n$, for every $j \in [k]$, $w_{jd} \in [2^{nR_{jd}}]$, and $z_j \in [2^{nC_{jd}}]$, let $U_j^n(w_{jd}, z_j | u_0^n)$ be drawn independently according to

$$\Pr \{U_j^n(w_{jd}, z_j | u_0^n) = u_j^n \mid U_0^n(w_0) = u_0^n\} = \prod_{t=1}^n p(u_{jt} | u_{0t}). \quad (12)$$

For every (w_1, \dots, w_k) , define $E(u_0^n, \mu_1, \dots, \mu_k)$ as the event where $U_0^n(w_0) = u_0^n$ and for every $j \in [k]$,

$$U_j^n(w_{jd}, \cdot | u_0^n) = \mu_j(\cdot), \quad (13)$$

where μ_j is a mapping from $[2^{nC_{jd}}]$ to \mathcal{U}_j^n . Let $\mathcal{A}(u_0^n, \mu_{[k]})$ be the set of all $z_{[k]} = (z_1, \dots, z_k)$ such that

$$(u_0^n, \mu_{[k]}(z_{[k]})) \in A_\delta^{(n)}(U_0, U_{[k]}), \quad (14)$$

where $\mu_{[k]}(z_{[k]}) = (\mu_1(z_1), \dots, \mu_k(z_k))$ and $A_\delta^{(n)}(U_0, U_{[k]})$ is the weakly typical set with respect to the distribution $p(u_0, u_{[k]})$. If $\mathcal{A}(u_0^n, \mu_{[k]})$ is empty, set $Z_j = 1$ for all $j \in [k]$. Otherwise, let the k -tuple $Z_{[k]} = (Z_1, \dots, Z_k)$ be the smallest element of $\mathcal{A}(u_0^n, \mu_{[k]})$ with respect to the lexicographical order. Finally, given $U_0^n(w_0) = u_0^n$ and $U_j^n(w_{jd}, Z_j | u_0^n) = u_j^n$, for each $w_{jj} \in [2^{nR_{jj}}]$, let $X_j^n(w_{jj} | u_0^n, u_j^n)$ be a random vector drawn independently according to

$$\begin{aligned} \Pr \{X_j^n(w_{jj} | u_0^n, u_j^n) = x_j^n \mid U_0^n(w_0) = u_0^n, U_j^n(w_{jd}, Z_j) = u_j^n\} \\ = \prod_{t=1}^n p(x_{jt} | u_{0t}, u_{jt}). \end{aligned}$$

We next describe the encoding and decoding processes.

Encoding. For every $j \in [k]$, encoder j sends the pair (w_{j0}, w_{jd}) to the CF. The CF sends $((w_{i0})_{i \neq j}, Z_j)$ back to encoder j . Encoder j , having access to $w_0 = (w_{j0})_j$ and Z_j , transmits $X_j^n(w_{jj} | U_0^n(w_0), U_j^n(w_{jd}, Z_j))$ over the channel.

Decoding. The decoder, upon receiving Y^n , maps Y^n to the unique k -tuple $\hat{W}_{[k]}$ such that

$$\begin{aligned} (U_0^n(\hat{W}_0), (U_j^n(\hat{W}_{jd}, \hat{Z}_j | U_0^n))_j, (X_j^n(\hat{W}_{jj} | U_0^n, U_j^n))_j, Y^n) \\ \in A_\epsilon^{(n)}(U_0, U_{[k]}, X_{[k]}, Y). \end{aligned} \quad (15)$$

If such a k -tuple does not exist, the decoder sets its output to the k -tuple $(1, 1, \dots, 1)$.

The analysis of the expected error probability for the proposed random code appears in Subsection VII-A.

V. CASE STUDY: 2-USER GAUSSIAN MAC

In this section, we study the network consisting of the 2-user Gaussian MAC with power constraints and a CF whose input link capacities are sufficiently large so that the CF has full access to the messages and output link capacities both equal C_{out} . We show that in this scenario, the benefit of cooperation extends beyond sum-capacity; that is, capacity metrics other than sum-capacity also exhibit an infinite slope at $C_{\text{out}} = 0$. In addition, we show that the behavior of these metrics (including sum-capacity) is bounded from below by a constant multiplied $\sqrt{C_{\text{out}}}$.

From Theorem 1, it follows that the capacity region of our network contains the set of all rate pairs (R_1, R_2) that satisfy

$$\begin{aligned} R_1 &\leq \max\{I(X_1; Y|U_0) - C_{1d}, I(X_1; Y|X_2, U_0) - \zeta\} + C_{10} \\ R_2 &\leq \max\{I(X_2; Y|U_0) - C_{2d}, I(X_2; Y|X_1, U_0) - \zeta\} + C_{20} \\ R_1 + R_2 &\leq I(X_1, X_2; Y|U_0) - \zeta + C_{10} + C_{20} \\ R_1 + R_2 &\leq I(X_1, X_2; Y) - \zeta \end{aligned}$$

for some nonnegative constants $C_{1d}, C_{2d} \leq C_{\text{out}}$,

$$C_{10} = C_{\text{out}} - C_{2d}$$

$$C_{20} = C_{\text{out}} - C_{1d},$$

and some distribution $p(u_0)p(x_1, x_2|u_0)$ that satisfies $\mathbb{E}[X_i^2] \leq P_i$ for $i \in \{1, 2\}$ and

$$\zeta := C_{1d} + C_{2d} - I(X_1; X_2|U_0) \geq 0.$$

By (1), the 2-user Gaussian MAC can be represented as

$$Y = X_1 + X_2 + Z,$$

where Z is independent of (X_1, X_2) , and is distributed as $Z \sim \mathcal{N}(0, N)$ for some noise variance $N > 0$. Let $U_0 \sim \mathcal{N}(0, 1)$, and (X'_1, X'_2) be a pair of random variables independent of U_0 and jointly distributed as $\mathcal{N}(\mu, \Sigma)$, where

$$\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma = \begin{pmatrix} 1 & \rho_0 \\ \rho_0 & 1 \end{pmatrix}$$

for some $\rho_0 \in [0, 1]$. Finally, for $i \in \{1, 2\}$, set

$$\frac{1}{\sqrt{P_i}} X_i = \rho_i X'_i + \sqrt{1 - \rho_i^2} U_0,$$

for some $\rho_i \in [0, 1]$. Calculating the region described above for the Gaussian MAC using the joint distribution of (U_0, X_1, X_2) and setting $\gamma_i = P_i/N$ for $i \in \{1, 2\}$ and $\bar{\gamma} = \sqrt{\gamma_1 \gamma_2}$, gives the set of all rate pairs (R_1, R_2) satisfying

$$\begin{aligned} R_1 &\leq \max \left\{ \frac{1}{2} \log \frac{1 + \rho_1^2 \gamma_1 + \rho_2^2 \gamma_2 + 2\rho_0 \rho_1 \rho_2 \bar{\gamma}}{1 + (1 - \rho_0^2) \rho_2^2 \gamma_2} - C_{1d}, \frac{1}{2} \log (1 + (1 - \rho_0^2) \rho_1^2 \gamma_1) - \zeta \right\} + C_{10} \\ R_2 &\leq \max \left\{ \frac{1}{2} \log \frac{1 + \rho_1^2 \gamma_1 + \rho_2^2 \gamma_2 + 2\rho_0 \rho_1 \rho_2 \bar{\gamma}}{1 + (1 - \rho_0^2) \rho_1^2 \gamma_1} - C_{2d}, \frac{1}{2} \log (1 + (1 - \rho_0^2) \rho_2^2 \gamma_2) - \zeta \right\} + C_{20} \end{aligned}$$

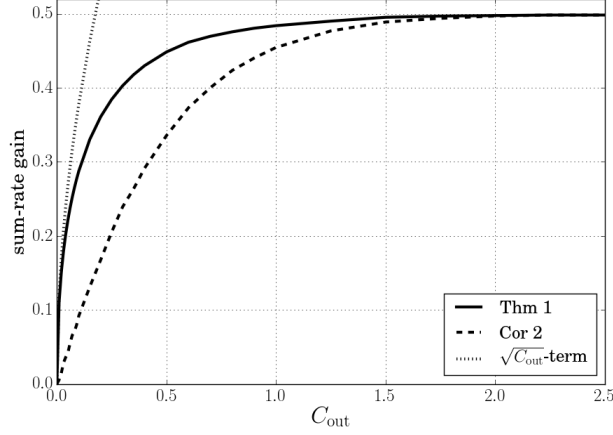


Figure 2. Plot of the achievable sum-rate gain given by Theorem 1 and Corollary 2 for Gaussian input distributions, and the $\sqrt{C_{\text{out}}}$ -term given in Proposition 7. Here $\gamma_1 = \gamma_2 = 100$.

and

$$R_1 + R_2 \leq \frac{1}{2} \log (1 + \rho_1^2 \gamma_1 + \rho_2^2 \gamma_2 + 2\rho_0 \rho_1 \rho_2 \bar{\gamma}) - \zeta + C_{10} + C_{20}$$

$$R_1 + R_2 \leq \frac{1}{2} \log \left(1 + \gamma_1 + \gamma_2 + 2(\rho_0 \rho_1 \rho_2 + \sqrt{(1 - \rho_1^2)(1 - \rho_2^2)}) \bar{\gamma} \right) - \zeta$$

for some $\rho_1, \rho_2 \in [0, 1]$, and $0 \leq \rho_0 \leq \sqrt{1 - 2^{-2(C_{1d} + C_{2d})}}$. Denote this region with $\mathcal{C}_{\text{ach}}(C_{\text{out}})$.

We next introduce a lower bound for the weighted version of the sum-capacity. Denote the capacity region of this network with $\mathcal{C}(C_{\text{out}})$. For every $\alpha \in [0, 1]$, define

$$C_\alpha(C_{\text{out}}) = \max_{(R_1, R_2) \in \mathcal{C}(C_{\text{out}})} (\alpha R_1 + (1 - \alpha) R_2)$$

Note that $C_\alpha(C_{\text{out}})$ is a generalization of the notion of sum-capacity where the weighted sum of the encoders' rates is considered. The main result of this section demonstrates that for small C_{out} , $C_\alpha(C_{\text{out}})$ is bounded from below by a constant times $\sqrt{C_{\text{out}}}$ when C_{out} is small. The proof is given in Subsection VII-E.

Proposition 7. *For the Gaussian MAC $Y = X_1 + X_2 + Z$ with $Z \sim \mathcal{N}(0, N)$ and input SNRs (γ_1, γ_2) , we have*

$$C_\alpha(C_{\text{out}}) - C_\alpha(0) \geq \frac{2\sqrt{\gamma_1 \gamma_2} \cdot \log e}{1 + \gamma_1 + \gamma_2} \cdot \min\{\alpha, 1 - \alpha\} \cdot \sqrt{C_{\text{out}}} + o(\sqrt{C_{\text{out}}}).$$

In particular, for every $\alpha \in (0, 1)$,

$$\left. \frac{dC_\alpha}{dC_{\text{out}}} \right|_{C_{\text{out}}=0^+} = \infty.$$

In Figure 2, using [19], we plot the sum-rate of the region $\mathcal{C}_{\text{ach}}(C_{\text{out}})$ and the forwarding inner bound (Corollary 2) for $\gamma_1 = \gamma_2 = 100$. We also plot the $\sqrt{C_{\text{out}}}$ -term in the lower bound given by Proposition 7. Notice that the forwarding inner bound provides a cooperation gain that is at most linear in C_{out} .

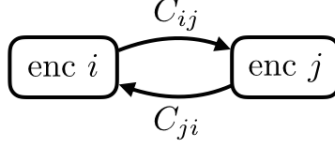


Figure 3. In k -user MAC with conferencing, for every $i, j \in [k]$, there are links of capacities C_{ij} and C_{ji} connecting encoders i and j .

VI. THE k -USER MAC WITH CONFERENCING ENCODERS

In this section, we extend Willems' conferencing encoders model [8] from the 2-user MAC to the k -user MAC and provide an outer bound on the capacity region.

Consider a k -user MAC where for every $i, j \in [k]$ (in this section, $i \neq j$ by assumption), there is a noiseless link of capacity $C_{ij} \geq 0$ going from encoder i to encoder j and a noiseless link of capacity $C_{ji} \geq 0$ going back (Figure 3). As in 2-user conferencing, the “conference” occurs over a finite number of rounds. In the first round, for every $i, j \in [k]$ with $C_{ij} > 0$, encoder i transmits some information to encoder j that is a function of its own message $w_i \in [2^{nR_i}]$. In each subsequent round, every encoder transmits information that is a function of its message and information it receives before that round. Once the conference is over, each encoder transmits its codeword over the k -user MAC.

We next define a $((2^{nR_1}, \dots, 2^{nR_k}), n, L)$ -code for the k -user MAC with an L -round $(C_{ij})_{i,j=1}^k$ -conference. For every $i, j \in [k]$ and $\ell \in [L]$, fix a set $\mathcal{V}_{ij}^{(\ell)}$ so that for every $i, j \in [k]$, $\sum_{\ell=1}^L \log |\mathcal{V}_{ij}^{(\ell)}| \leq nC_{ij}$. Here $\mathcal{V}_{ij}^{(\ell)}$ represents the alphabet of the symbol encoder i sends to encoder j in round ℓ of the conference. For every $\ell \in [L]$, define $\mathcal{V}_{ij}^\ell = \prod_{\ell'=1}^{\ell} \mathcal{V}_{ij}^{(\ell')}$. For $j \in [k]$, encoder j is represented by the collection of functions $(f_j, (h_{ji}^{(\ell)})_{i,\ell})$ where

$$\begin{aligned} f_j &: [2^{nR_j}] \times \prod_{i:i \neq j} \mathcal{V}_{ij}^L \rightarrow \mathcal{X}_j^n \\ h_{ji}^{(\ell)} &: [2^{nR_j}] \times \prod_{i':i' \neq j} \mathcal{V}_{i'j}^{\ell-1} \rightarrow \mathcal{V}_{ji}^{(\ell)} \end{aligned}$$

The decoder is a mapping $g: \mathcal{Y}^n \rightarrow \prod_{j=1}^k [2^{nR_j}]$. The definitions of cost constraints, achievable rate vectors, and the capacity region are similar to those given in Section II.

The next result compares the capacity region of a MAC with cooperation under the conferencing and CF models. The proof is given in Subsection VII-F.

Proposition 8. *The capacity region of a MAC with an L -round $(C_{ij})_{i,j=1}^k$ -conference is a subset of the capacity region of the same MAC with an L -round $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$ -CF cooperation if for all $j \in [k]$,*

$$C_{\text{in}}^j \geq \sum_{i:i \neq j} C_{ji} \quad \text{and} \quad C_{\text{out}}^j \geq \sum_{i:i \neq j} C_{ij}.$$

Similarly, for every L , the capacity region of a MAC with L -round $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$ -CF cooperation is a subset of the capacity region of the same MAC with a single-round $(C_{ij})_{i,j=1}^k$ -conference if for all $i, j \in [k]$, $C_{ij} \geq C_{\text{in}}^i$.

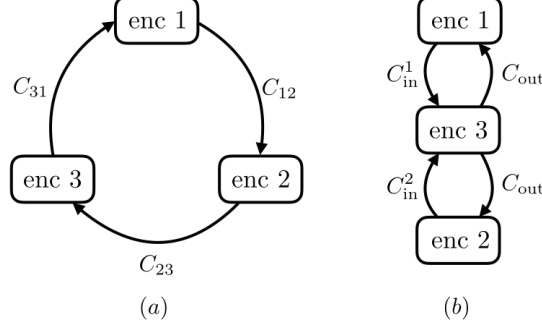


Figure 4. (a) The conferencing structure studied in [17]. (b) An example of a structure where allowing two conferencing rounds leads to a substantial gain over a single round.

Combining the first part of Proposition 8 with the outer bound from Proposition 5 results in the next corollary, which holds regardless of the number of conferencing rounds.

Corollary 9 (Conferencing Outer Bound). *The capacity region of a MAC with a $(C_{ij})_{i,j=1}^k$ -conference is a subset of the set of all rate vectors (R_1, \dots, R_k) that for some distribution $p \in \mathcal{P}_{\text{ind}}(\mathcal{U}_0, \mathcal{X}_{[k]})$, satisfy*

$$\sum_{j \in S} R_j \leq I(X_S; Y | U_0, X_{S^c}) + \sum_{j \in S} \sum_{i \neq j} C_{ji} \quad \forall \emptyset \neq S \subseteq [k]$$

$$\sum_{j \in [k]} R_j \leq I(X_{[k]}; Y).$$

While k -user conferencing is a direct extension of 2-user conferencing, there is nonetheless an important difference when $k \geq 3$. While a single conferencing round suffices to achieve the capacity region in the 2-user case [8], the same is not true when $k \geq 3$, as we next see.

A special case of this model for the 3-user Gaussian MAC, depicted in Figure 4(a), is studied in [17]. While the achievability scheme in [17] uses two conferencing rounds, the magnitude of the gain resulting from using an additional conferencing round is not clear. Here, using the idea of a cooperation facilitator, we consider an alternative shown in Figure 4(b), where we show the possibility of a large cooperation gain when conferencing occurs in two rounds rather than one. Consider a 3-user MAC with conferencing. Fix positive constants C_{in}^1 and C_{in}^2 . Let $C_{13} = C_{in}^1$, $C_{23} = C_{in}^2$, $C_{31} = C_{32} = C_{out}$ for $C_{out} \in \mathbb{R}_{\geq 0}$, and $C_{12} = C_{21} = 0$. Let $\mathcal{C}_1(C_{out})$ and $\mathcal{C}_2(C_{out})$ denote the capacity region of this network with one and two rounds of conferencing, respectively. For each $L \in \{1, 2\}$, define the function $g_L(C_{out})$ as

$$g_L(C_{out}) = \max_{(R_1, R_2, 0) \in \mathcal{C}_L(C_{out})} (R_1 + R_2).$$

Note that when $L = 1$, we have $g_1(C_{out}) = g_1(0)$ for all $C_{out} \geq 0$, since no cooperation is possible when encoder 3 is transmitting at rate zero. On the other hand, we next show that at least for some MACs, $g_2'(0) = \infty$; that is, g_2 has an infinite slope at $C_{out} = 0$. Note that

$$g_2(0) = g_1(0) = \max_{p(x_1)p(x_2), x_3} I(X_1, X_2; Y | X_3 = x_3).$$

Suppose x_3^* satisfies

$$\max_{p(x_1)p(x_2), x_3} I(X_1, X_2; Y | X_3 = x_3) = \max_{p(x_1)p(x_2)} I(X_1, X_2; Y | X_3 = x_3^*).$$

If the MAC $(\mathcal{X}_1 \times \mathcal{X}_2, p(y|x_1, x_2, x_3^*), \mathcal{Y})$ is in $\mathcal{C}^*(\mathcal{X}_1 \times \mathcal{X}_2, \mathcal{Y})$, then by Theorem 3, we have $g_2'(0) = \infty$. Since g_1 is constant for all C_{out} , while g_2 has an infinite slope at $C_{\text{out}} = 0$, and $g_1(0) = g_2(0)$, the two-round conferencing region is strictly larger than the single-round conferencing region. Using the same technique, we can show a similar result for any $k \geq 3$; that is, there exist k -user MACs where the two-round conferencing region strictly contains the single-round region.

VII. PROOFS

A. Theorem 1 (Inner bound)

Fix $\eta > 0$, and choose a distribution $p(u_0, u_{[k]}, x_{[k]})$ on $\mathcal{U}_0 \times \mathcal{U}_{[k]} \times \mathcal{X}_{[k]}$ of the form

$$p(u_0) \cdot \prod_{i \in S_d^c} p(u_i | u_0) \cdot p(u_{S_d} | u_0, u_{S_d^c}) \cdot \prod_{j \in [k]} p(x_j | u_0, u_j),$$

that satisfies the dependence constraints

$$\zeta_S := \sum_{j \in S} C_{jd} - \sum_{j \in S} H(U_j | U_0) + H(U_S | U_0, U_{S_d^c}) > 0 \quad \forall \emptyset \subsetneq S \subseteq S_d,$$

and cost constraints

$$\mathbb{E}[b_j(X_j)] \leq B_j - \eta \quad \forall j \in [k]. \quad (16)$$

Let (w_1, \dots, w_k) denote the transmitted k -tuple of messages and $(\hat{W}_1, \dots, \hat{W}_k)$ denote the output of the decoder. To simplify notation, denote

$$U_0^n(w_0), U_j^n(w_{jd}, Z_j | U_0^n), X_j^n(w_{jj} | U_0^n, U_j^n)$$

with U_0^n , U_j^n , and X_j^n , respectively. Similarly, define \hat{U}_0^n , \hat{U}_j^n , and \hat{X}_j^n as

$$U_0^n(\hat{W}_0), U_j^n(\hat{W}_{jd}, Z_j | U_0^n), X_j^n(\hat{W}_{jj} | U_0^n, U_j^n).$$

Here \hat{W}_0 , \hat{W}_{jd} , and \hat{W}_{jj} are defined in terms of $(\hat{W}_j)_j$ similar to the definitions of w_0 , w_{jd} , and w_{jj} in Section IV. Let Y^n denote the channel output when $X_{[k]}^n$ is transmitted. Then the joint distribution of $(U_0^n, U_{[k]}^n, X_{[k]}^n, Y^n)$ is given by

$$p_{\text{code}}(u_0^n, u_{[k]}^n, x_{[k]}^n, y^n) = p(u_0^n) p_{\text{code}}(u_{[k]}^n | u_0^n) p(x_{[k]}^n | u_0^n, u_{[k]}^n) p(y^n | x_{[k]}^n),$$

where

$$p_{\text{code}}(u_{[k]}^n | u_0^n) = \sum_{\mu_{[k]}} p(\mu_1 | u_0^n) \dots p(\mu_k | u_0^n) p(u_{[k]}^n | u_0^n, \mu_{[k]})$$

and $p(\mu_j | u_0^n)$ and $p(u_{[k]}^n | u_0^n, \mu_{[k]})$ are calculated according to

$$p(\mu_j | u_0^n) = \prod_{z_j \in [2^{n C_{jd}}]} p(\mu_j(z_j) | u_0^n),$$

and

$$p(u_{[k]}^n | u_0^n, \mu_{[k]}) = \sum_{z_{[k]}} p(z_{[k]} | u_0^n, \mu_{[k]}) \prod_{j=1}^k \mathbf{1}\{\mu_j(z_j) = u_j^n\}.$$

Define the distribution $p_{\text{ind}}(u_0^n, u_{[k]}^n, x_{[k]}^n, y^n)$ as

$$p_{\text{ind}}(u_0^n, u_{[k]}^n, x_{[k]}^n, y^n) = p(u_0^n) p(x_{[k]}^n | u_0^n, u_{[k]}^n) p(y^n | x_{[k]}^n) \prod_{j=1}^k p(u_j^n | u_0^n),$$

which is the joint input-output distribution if independent codewords are transmitted. We next mention some results regarding weakly typical sets that are required for our error analysis.

For any $S \subseteq [k]$, let $A_\delta^{(n)}(U_0, U_S)$ denote the weakly typical set with respect to the distribution $p(u_0, u_S)$, a marginal of $p(u_0, u_{[k]})$. In addition, for every $(u_0^n, u_S^n) \in A_\delta^{(n)}(U_0, U_S)$, let $A_\delta^{(n)}(u_0^n, u_S^n)$ be the set of all $u_{S^c}^n$ such that

$$(u_0^n, u_{[k]}^n) \in A_\delta^{(n)}(U_0, U_{[k]}).$$

Similarly, let $A_\epsilon^{(n)}(U_0, U_{[k]}, X_{[k]}, Y)$ be the weakly typical set with respect to the distribution $p(u_0, u_{[k]}, x_{[k]}, y)$, where $p(y | x_{[k]})$ is given by the channel definition. For subsets $S, T \subseteq [k]$, define $A_\epsilon^{(n)}(U_0, U_S, X_T, Y)$ and $A_\epsilon^{(n)}(u_0^n, u_S^n, x_T^n, y^n)$ accordingly. If $(u_0^n, u_S^n, x_T^n, y^n) \in A_\epsilon^{(n)}(U_0, U_S, X_T, Y)$, we have [13, p. 523]

$$\log |A_\epsilon^{(n)}(u_0^n, u_S^n, x_T^n, y^n)| \leq n(H(U_{S^c}, X_{T^c} | U_0, U_S, X_T, Y) + 2\epsilon). \quad (17)$$

Finally, under fairly general conditions described in Appendix B,⁴ there exists an increasing function $I : \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}$ such that if $(U_0^n, U_{[k]}^n, X_{[k]}^n, Y^n)$ consists of n i.i.d. copies of $(U_0, U_{[k]}, X_{[k]}, Y)$ distributed according to $p(u_0, u_{[k]}, x_{[k]}, y)$, then

$$\Pr \left\{ (U_0^n, U_{[k]}^n, X_{[k]}^n, Y^n) \in A_\epsilon^{(n)}(U_0, U_{[k]}, X_{[k]}, Y) \right\} \geq 1 - 2^{-nI(\epsilon)}. \quad (18)$$

Fix any such function I .

We next study the relationship between p_{code} and p_{ind} . Our first lemma provides an upper bound for p_{code} in terms of p_{ind} .

Lemma 10. *For every nonempty $S \subseteq [k]$ and all (u_0^n, u_S^n) ,*

$$\frac{1}{n} \log \frac{p_{\text{code}}(u_S^n | u_0^n)}{p_{\text{ind}}(u_S^n | u_0^n)} \leq nC_{Sd},$$

where $C_{Sd} = \sum_{j \in S} C_{jd}$.

Proof: Recall

$$p_{\text{code}}(u_S^n | u_0^n) = \sum_{\mu_{[k]}} p(u_S^n | u_0^n, \mu_{[k]}) \prod_{j \in [k]} p(\mu_j | u_0^n).$$

To bound $p_{\text{code}}(u_S^n | u_0^n)$, note that

$$p(u_S^n | u_0^n, \mu_{[k]}) \leq \prod_{j \in S} \mathbf{1}\{\mu_j^{-1}(u_j^n) \neq \emptyset\},$$

⁴Distributions that satisfy these conditions include any distribution with finite support and the Gaussian distribution.

where

$$\mu_j^{-1}(u_j^n) = \{z_j \in [2^{n C_{jd}}] : \mu_j(z_j) = u_j^n\}.$$

Now for every $j \in S$,

$$\begin{aligned} \sum_{\mu_j} p(\mu_j | u_0^n) \mathbf{1}\{\mu_j^{-1}(u_j^n) \neq \emptyset\} &= \Pr \{ \exists z_j : U_j^n(z_j) = u_j^n | U_0^n = u_0^n \} \\ &\leq 2^{n C_{jd}} p(u_j^n | u_0^n). \end{aligned}$$

Thus

$$\begin{aligned} p_{\text{code}}(u_S^n | u_0^n) &\leq \sum_{\mu_S} \prod_{j \in S} p(\mu_j | u_0^n) \mathbf{1}\{\mu_j^{-1}(u_j^n) \neq \emptyset\} \\ &= \prod_{j \in S} \left(\sum_{\mu_j} p(\mu_j | u_0^n) \mathbf{1}\{\mu_j^{-1}(u_j^n) \neq \emptyset\} \right) \\ &\leq 2^{n \sum_{j \in S} C_{jd}} p_{\text{ind}}(u_S^n | u_0^n). \end{aligned}$$

■

Our second lemma provides an upper bound for $p_{\text{ind}}(u_S^n | u_0^n)$ when (u_0^n, u_S^n) is typical.

Lemma 11. *For all nonempty $S_d^c \subseteq S \subseteq [k]$ and $(u_0^n, u_S^n) \in A_\delta^{(n)}(U_0, U_S)$,*

$$\frac{1}{n} \log \frac{p_{\text{ind}}(u_S^n | u_0^n)}{p(u_S^n | u_0^n)} \leq - \sum_{j \in S \cap S_d} H(U_j | U_0) + H(U_{S \cap S_d} | U_0, U_{S_d^c}) + 2(|S \cap S_d| + 1)\delta.$$

Proof: Recall that

$$p(u_{[k]}^n | u_0^n) = p(u_{S_d}^n | u_0^n, u_{S_d^c}^n) \prod_{j \in S_d^c} p(u_j^n | u_0^n).$$

Thus for all $S \supseteq S_d^c$, we have

$$p(u_S^n | u_0^n) = p(u_{S \cap S_d}^n | u_0^n, u_{S_d^c}^n) \prod_{j \in S_d^c} p(u_j^n | u_0^n).$$

Therefore,

$$\begin{aligned} \frac{p_{\text{ind}}(u_S^n | u_0^n)}{p(u_S^n | u_0^n)} &= \frac{p_{\text{ind}}(u_{S \cap S_d}^n | u_0^n)}{p(u_{S \cap S_d}^n | u_0^n, u_{S_d^c}^n)} \\ &= \frac{\prod_{j \in S \cap S_d} p(u_j^n | u_0^n)}{p(u_{S \cap S_d}^n | u_0^n, u_{S_d^c}^n)}. \end{aligned}$$

The proof now follows from the definition of $A_\delta^{(n)}(U_0, U_S)$. ■

Combining the previous two lemmas results in the next corollary, which we use in our error analysis.

Corollary 12. *For every nonempty S satisfying $S_d^c \subseteq S \subseteq [k]$ and all $(u_0^n, u_S^n) \in A_\delta^{(n)}(U_0, U_S)$,*

$$\frac{1}{n} \log \frac{p_{\text{code}}(u_S^n | u_0^n)}{p(u_S^n | u_0^n)} \leq \zeta_{S \cap S_d} + 2(|S \cap S_d| + 1)\delta.$$

Let \mathcal{E} denote the event where either the output of an encoder does not satisfy the corresponding cost constraint, or the output of the decoder differs from the transmitted k -tuple of messages; that is $(\hat{W}_j)_{j=1}^k \neq (w_j)_{j=1}^k$. Denote

the former event with $\mathcal{E}_{\text{cost}}$ and the latter event with \mathcal{E}_{dec} . When \mathcal{E}_{dec} occurs, it is either the case that $(w_j)_{j=1}^k$ does not satisfy (15) (denote this event with \mathcal{E}_{typ}), or that there is another k -tuple, $(\hat{W}_j)_{j=1}^k \neq (w_j)_{j=1}^k$, that also satisfies (15). If the latter event occurs, we either have $\hat{W}_0 \neq w_0$ (denote event with $\mathcal{E}_{\emptyset, \emptyset}$), or $\hat{W}_0 = w_0$. When $\hat{W}_0 = w_0$, define the subsets $S, T \subseteq [k]$ as

$$S = \{j : \hat{W}_{jd} \neq w_{jd}\}$$

$$T = \{j : \hat{W}_{jj} \neq w_{jj}\}.$$

Now for every pair of subsets $S, T \subseteq [k]$ such that $S \cup T \neq \emptyset$, define $\mathcal{E}_{S,T}$ as the event where there exists a $(\hat{W}_j)_{j=1}^k$ that satisfies (15), $\hat{W}_0 = w_0$, $\hat{W}_{jd} \neq w_{jd}$ if and only if $j \in S$, and $\hat{W}_{jj} \neq w_{jj}$ if and only if $j \in T$. Thus we may write

$$\mathcal{E} \subseteq \mathcal{E}_{\text{cost}} \cup \mathcal{E}_{\text{typ}} \cup \bigcup_{S, T \subseteq [k]} \mathcal{E}_{S,T}.$$

The union over all $\mathcal{E}_{S,T}$ also contains the event $\mathcal{E}_{\emptyset, \emptyset}$. By the union bound,

$$\Pr(\mathcal{E}) \leq \Pr(\mathcal{E}_{\text{cost}}) + \Pr(\mathcal{E}_{\text{typ}}) + \sum_{S, T \subseteq [k]} \Pr(\mathcal{E}_{S,T}).$$

Thus to find a set of achievable rates for our random code design, it suffices to find conditions under which $\Pr(\mathcal{E}_{\text{cost}})$, $\Pr(\mathcal{E}_{\text{typ}})$, and each $\Pr(\mathcal{E}_{S,T})$ go to zero as $n \rightarrow \infty$.

We begin our analysis with the event $\mathcal{E}_{\text{cost}}$. For $j \in [k]$, let $\mathcal{E}_{\text{cost}}^j$ denote the event where the codeword $X_j^n(w_{jj}|U_0^n(w_0), U_j^n(w_{jd}, Z_j))$ does not satisfy the cost constraint of encoder j . We have

$$\begin{aligned} \Pr(\mathcal{E}_{\text{cost}}^j) &= \Pr\left\{\frac{1}{n} \sum_{t=1}^n b_j(X_{jt}(w_{jj}|U_0^n(w_0), U_j^n(w_{jd}, Z_j))) > B_j\right\} \\ &= \sum_{z_j} \Pr\{Z_j = z_j\} \Pr\left\{\frac{1}{n} \sum_{t=1}^n b_j(X_{jt}(w_{jj}|U_0^n(w_0), U_j^n(w_{jd}, z_j))) > B_j\right\}. \end{aligned}$$

Since for all z_j , by the AEP,

$$\Pr\left\{\frac{1}{n} \sum_{t=1}^n b_j(X_{jt}(w_{jj}|U_0^n(w_0), U_j^n(w_{jd}, z_j))) > B_j\right\} \rightarrow 0$$

as $n \rightarrow \infty$, it follows that $\Pr(\mathcal{E}_{\text{cost}}^j) \rightarrow 0$. Applying the union bound now implies

$$\Pr(\mathcal{E}_{\text{cost}}) \leq \sum_{j \in [k]} \Pr(\mathcal{E}_{\text{cost}}^j) \rightarrow 0.$$

We next consider the event \mathcal{E}_{typ} . Define \mathcal{E}_{enc} as the event where

$$(U_0^n, U_{[k]}^n) \notin A_\delta^{(n)}(U_0, U_{[k]})$$

and note that \mathcal{E}_{typ} is the event where

$$(U_0^n, U_{[k]}^n, X_{[k]}^n, Y^n) \notin A_\epsilon^{(n)}(U_0, U_{[k]}, X_{[k]}, Y).$$

The event \mathcal{E}_{enc} occurs if and only if $\mathcal{A}(U_0^n, U_{[k]}^n(\cdot))$ (defined in Section IV) is empty. Thus

$$\Pr(\mathcal{E}_{\text{enc}}) = \Pr\{\mathcal{A}(U_0^n, U_{[k]}^n(\cdot)) = \emptyset\}.$$

If $S_d = \emptyset$, $\Pr(\mathcal{E}_{\text{enc}})$ goes to zero by the AEP since in this case $p_{\text{code}}(u_{[k]}^n | u_0^n) = p(u_{[k]}^n | u_0^n)$. Otherwise, recall that for every nonempty $S \subseteq S_d$, ζ_S is defined as

$$\zeta_S = \sum_{j \in S} C_{jd} - \sum_{j \in S} H(U_j | U_0) + H(U_S | U_0, U_{S_d^c}).$$

From the multivariate covering lemma (Appendix A), it follows that $\Pr(\mathcal{E}_{\text{enc}}) \rightarrow 0$ if for all nonempty $S \subseteq S_d$,

$$\zeta_S > (8|S_d| - 2|S| + 10)\delta. \quad (19)$$

Next we find an upper bound for $\Pr(\mathcal{E}_{\text{typ}} \setminus \mathcal{E}_{\text{enc}})$. Let $B^{(n)}$ be the set of all $(u_0^n, u_{[k]}^n, x_{[k]}^n, y^n)$ such that $(u_0^n, u_{[k]}^n) \in A_\delta^{(n)}$ but $(u_0^n, u_{[k]}^n, x_{[k]}^n, y^n) \notin A_\epsilon^{(n)}$. Then

$$\begin{aligned} \Pr(\mathcal{E}_{\text{typ}} \setminus \mathcal{E}_{\text{enc}}) &= \sum_{B^{(n)}} p(u_0^n) p_{\text{code}}(u_{[k]}^n | u_0^n) p(x_{[k]}^n | u_0^n, u_{[k]}^n) p(y^n | x_{[k]}^n) \\ &\stackrel{(a)}{\leq} 2^{n(\zeta_{S_d} + 2(|S_d| + 1)\delta)} \sum_{B^{(n)}} p(u_0^n, u_{[k]}^n, x_{[k]}^n, y^n) \\ &\stackrel{(b)}{\leq} 2^{n(\zeta_{S_d} + 2(|S_d| + 1)\delta)} \Pr\{(A_\epsilon^{(n)})^c\} \stackrel{(c)}{\leq} 2^{n(\zeta_{S_d} + 2(|S_d| + 1)\delta - I(\epsilon))}, \end{aligned}$$

where (a) follows from Corollary 12, (b) holds since $B^{(n)} \subseteq (A_\epsilon^{(n)})^c$, and (c) follows from the definition of $I(\epsilon)$ given by (18). Thus $\Pr(\mathcal{E}_{\text{typ}} \setminus \mathcal{E}_{\text{enc}}) \rightarrow 0$ if

$$\zeta_{S_d} < I(\epsilon) - 2(|S_d| + 1)\delta. \quad (20)$$

Therefore, if (19) and (20) both hold, then $\Pr(\mathcal{E}_{\text{typ}}) \rightarrow 0$ since

$$\Pr(\mathcal{E}_{\text{typ}}) \leq \Pr(\mathcal{E}_{\text{enc}} \cup \mathcal{E}_{\text{typ}}) = \Pr(\mathcal{E}_{\text{enc}}) + \Pr(\mathcal{E}_{\text{typ}} \setminus \mathcal{E}_{\text{enc}}).$$

We next study $\mathcal{E}_{\emptyset, \emptyset}$, which is the event where there exists a k -tuple $(\hat{W}_j)_j$ that satisfies (15) but $\hat{W}_0 \neq w_0$. If this event occurs, then $(\hat{U}_0^n, \hat{U}_{[k]}^n, \hat{X}_{[k]}^n)$ and Y^n are independent. By the union bound,

$$\Pr(\mathcal{E}_{\emptyset, \emptyset}) \leq 2^{n \sum_{j=1}^k R_j} \sum_{A_\epsilon^{(n)}} p_{\text{code}}(u_0^n, u_{[k]}^n, x_{[k]}^n) p_{\text{code}}(y^n).$$

We rewrite the sum in the above inequality as

$$\sum_{A_\epsilon^{(n)}(Y)} p_{\text{code}}(y^n) \sum_{A_\epsilon^{(n)}(y^n)} p_{\text{code}}(u_0^n, u_{[k]}^n, x_{[k]}^n),$$

Using Corollary 12, we upper bound the inner sum by

$$\begin{aligned} &\sum_{A_\epsilon^{(n)}(y^n)} 2^{n(\zeta_{S_d} + 2(|S_d| + 1)\delta)} p(u_0^n, u_{[k]}^n, x_{[k]}^n) \\ &\stackrel{(*)}{\leq} 2^{n(H(U_0, U_{[k]}, X_{[k]} | Y) + 2\epsilon)} 2^{n(\zeta_{S_d} + 2(|S_d| + 1)\delta)} 2^{-n(H(U_0, U_{[k]}, X_{[k]}) + \epsilon)}, \end{aligned}$$

where $(*)$ follows from (17). This implies $\Pr(\mathcal{E}_{\emptyset, \emptyset}) \rightarrow 0$ if

$$\sum_{j=1}^k R_j < I(X_{[k]}; Y) - \zeta_{S_d} - 2(|S_d| + 1)\delta - 3\epsilon.$$

Next, let $S, T \subseteq [k]$ be sets such that $S \cup T \neq \emptyset$ and consider the event $\mathcal{E}_{S,T}$. Recall that this is the event where there exists a k -tuple $(\hat{W}_j)_j$ that satisfies (15) and $\hat{W}_0 = w_0$, $\hat{W}_{jd} \neq w_{jd}$ if and only if $j \in S$, and $\hat{W}_{jj} \neq w_{jj}$ if and only if $j \in T$. For every $A \subseteq S$ and $B \subseteq S^c$, let $\mathcal{E}_{S,T}^{A,B} \subseteq \mathcal{E}_{S,T}$ be the event where there exists a k -tuple $(\hat{W}_j)_j$ that satisfies

$$\begin{aligned} & \left(U_0^n(w_0), (U_j^n(\hat{W}_{jd}, \hat{Z}_j | U_0^n))_{j \in A}, (U_j^n(w_{jd}, \hat{Z}_j | U_0^n))_{j \in B}, \right. \\ & \left. (X_j^n(\hat{W}_{jj} | U_0^n, \hat{U}_j^n))_{j \in A \cup (B \cap T)}, (X_j^n(w_{jj} | U_0^n, \hat{U}_j^n))_{j \in B \setminus T}, Y^n \right) \in A_\epsilon^{(n)} \end{aligned} \quad (21)$$

and $\hat{W}_0 = w_0$, $\hat{W}_{jd} \neq w_{jd}$ if and only if $j \in S$, and $\hat{W}_{jj} \neq w_{jj}$ if and only if $j \in T$. If $\mathcal{E}_{S,T}$ occurs, then so does $\mathcal{E}_{S,T}^{A,B}$ for every $A \subseteq S$ and $B \subseteq S^c$. Thus

$$\mathcal{E}_{S,T} \subseteq \bigcap_{A,B} \mathcal{E}_{S,T}^{A,B}.$$

This implies

$$\Pr(\mathcal{E}_{S,T}) \leq \min_{A,B} \Pr(\mathcal{E}_{S,T}^{A,B}). \quad (22)$$

Therefore, to bound $\Pr(\mathcal{E}_{S,T})$, we find an upper bound on $\Pr(\mathcal{E}_{S,T}^{A,B})$ for any $A \subseteq S$ and $B \subseteq S^c$ such that $A \cup (B \cap T) \neq \emptyset$. This is the key difference between our error analysis here and the error analysis for the 2-user MAC with transmitter cooperation presented in [1]. For independent distributions, using the constraint that subsets of typical codewords are also typical does not lead to a larger region; the same may not be true when dealing with dependent distributions. That being said, to include all independent random variables in our error analysis, instead of calculating the minimum in (22) over all $A \subseteq S$ and $B \subseteq S^c$, we limit ourselves to subsets A and B that satisfy

$$S \cap S_d^c \subseteq A \subseteq S$$

$$S^c \cap S_d^c \subseteq B \subseteq S^c,$$

since all the random vectors $(U_j^n)_{j \in S_d^c}$ are independent given U_0^n . Choose any such A and B . Note that for every $j \in A \cup (B \cap T)$, either $\hat{W}_{jd} \neq w_{jd}$ or $\hat{W}_{jj} \neq w_{jj}$. In addition, in (21),

$$\begin{aligned} & \left((U_j^n(\hat{W}_{jd}, \hat{Z}_j | U_0^n))_{j \in A}, (U_j^n(w_{jd}, \hat{Z}_j | U_0^n))_{j \in B}, \right. \\ & \left. (X_j^n(\hat{W}_{jj} | U_0^n, U_j^n))_{j \in A \cup (B \cap T)}, (X_j^n(w_{jj} | U_0^n, U_j^n))_{j \in B \setminus T} \right) \end{aligned}$$

is independent of Y^n given

$$\left(U_0^n(w_0), (U_j^n(w_{jd}, \cdot | U_0^n))_{j \in S^c}, (X_j^n(w_{jj} | U_0^n, U_j^n(\cdot)))_{j \in S^c \setminus T} \right).$$

Therefore, by the union bound, $\Pr(\mathcal{E}_{S,T}^{A,B})$ is bounded from above by

$$\begin{aligned} & 2^n \left(\sum_{j \in A} R_{jd} + \sum_{j \in A \cup (B \cap T)} R_{jj} \right) \\ & \times \sum_{A_\epsilon^{(n)}} p(x_{A \cup (B \cap T)}^n | u_0^n, u_{A \cup (B \cap T)}^n) \\ & \times \sum_{\mu_{A \cup S^c}, \chi_{S^c \setminus T}} p(u_0^n, \mu_{S^c}, \chi_{S^c \setminus T}, y^n) p(\mu_A | u_0^n) p(u_{A \cup B}^n, x_{B \setminus T}^n | u_0^n, \mu_{A \cup S^c}, \chi_{S^c \setminus T}), \end{aligned} \quad (23)$$

where the inner sum is over all mappings $\mu_j : [2^{nC_{jd}}] \rightarrow \mathcal{U}_j^n$ for $j \in A \cup S^c$ and $\chi_j : [2^{nC_{jd}}] \rightarrow \mathcal{X}_j^n$ for $j \in S^c \setminus T$.

The distribution $p(u_0^n, \mu_{S^c}, \chi_{S^c \setminus T}, y^n)$ is a marginal of $p(u_0^n, \mu_{[k]}, \chi_{[k]}, y^n)$, which is defined as

$$p(u_0^n, \mu_{[k]}, \chi_{[k]}, y^n) = p(u_0^n, \mu_{[k]})p(\chi_{[k]}|u_0^n, \mu_{[k]})p(y^n|u_0^n, \mu_{[k]}, \chi_{[k]}),$$

where

$$\begin{aligned} p(\chi_{[k]}|u_0^n, \mu_{[k]}) &= \prod_{j \in [k]} p(\chi_j|u_0^n, \mu_j) \\ &= \prod_{j \in [k]} \prod_{z_j \in [2^{nC_{jd}}]} p(\chi_j(z_j)|u_0^n, \mu_j(z_j)), \end{aligned}$$

and

$$p(y^n|u_0^n, \mu_{[k]}, \chi_{[k]}) = \sum_{z_{[k]}} p(z_{[k]}|u_0^n, \mu_{[k]})p(y^n|\chi_{[k]}(z_{[k]})).$$

We have

$$\begin{aligned} &p(u_{A \cup B}^n, x_{B \setminus T}^n | u_0^n, \mu_{A \cup S^c}, \chi_{S^c \setminus T}) \\ &\leq \mathbf{1} \left\{ \exists (z_j)_{j \in B} \in \prod_{j \in B} [2^{nC_{jd}}] : (\forall j \in B : \mu_j(z_j) = u_j^n) \wedge (\forall j \in B \setminus T : \chi_j(z_j) = x_j^n) \right\} \\ &\quad \times \mathbf{1} \left\{ \exists (z_j)_{j \in A} \in \prod_{j \in A} [2^{nC_{jd}}] : (\forall j \in A : \mu_j(z_j) = u_j^n) \right\}. \end{aligned} \tag{24}$$

We can thus upper bound the inner sum in (23) as a product of the sums

$$\begin{aligned} &\sum_{\mu_{S^c}, \chi_{S^c \setminus T}} p(u_0^n, \mu_{S^c}, \chi_{S^c \setminus T}, y^n) \\ &\quad \times \mathbf{1} \left\{ \exists (z_j)_{j \in B} \in \prod_{j \in B} [2^{nC_{jd}}] : (\forall j \in B : \mu_j(z_j) = u_j^n) \wedge (\forall j \in B \setminus T : \chi_j(z_j) = x_j^n) \right\} \end{aligned}$$

and

$$\sum_{\mu_A} p(\mu_A | u_0^n) \mathbf{1} \left\{ \exists (z_j)_{j \in A} \in \prod_{j \in A} [2^{nC_{jd}}] : \forall j \in A, \mu_j(z_j) = u_j^n \right\}.$$

We first find an upper bound for the first sum. Define the distribution $\tilde{p}(u_0^n, u_{[k]}^n, x_{[k]}^n, y^n)$ as

$$\tilde{p}(u_0^n, u_{[k]}^n, x_{[k]}^n, y^n) = \sum_{\mu_{[k]}, \chi_{[k]}} p(u_0^n, \mu_{[k]}, \chi_{[k]}, y^n) \prod_{j=1}^k \mathbf{1} \{ \mu_j(1) = u_j^n, \chi_j(1) = x_j^n \}.$$

The following argument demonstrates that $\tilde{p}(u_0^n, u_{[k]}^n, x_{[k]}^n) = p_{\text{ind}}(u_0^n, u_{[k]}^n, x_{[k]}^n)$,

$$\begin{aligned} \tilde{p}(u_0^n, u_{[k]}^n, x_{[k]}^n) &= \sum_{y^n} \tilde{p}(u_0^n, u_{[k]}^n, x_{[k]}^n, y^n) \\ &= \sum_{\mu_{[k]}, \chi_{[k]}} p(u_0^n, \mu_{[k]}, \chi_{[k]}) \prod_{j=1}^k \mathbf{1} \{ \mu_j(1) = u_j^n, \chi_j(1) = x_j^n \} \\ &= p(u_0^n) \prod_{j=1}^k \sum_{\mu_j, \chi_j} p(\mu_j, \chi_j | u_0^n) \mathbf{1} \{ \mu_j(1) = u_j^n, \chi_j(1) = x_j^n \} \\ &= p_{\text{ind}}(u_0^n, u_{[k]}^n, x_{[k]}^n). \end{aligned} \tag{25}$$

For every $\mathbf{z} = (z_j)_{j \in B}$, where $z_j \in [2^{nC_{jd}}]$ for all $j \in B$, let $E_{\mathbf{z}}$ denote the event where for all $j \in B$, $U_j^n(w_{jd}, z_j | U_0^n) = u_j^n$, and for all $j \in B \setminus T$, $X_j^n(w_{jj} | U_0^n, U_j^n) = x_j^n$. Then

$$\begin{aligned}
& \sum_{\mu_{S^c}, \chi_{S^c \setminus T}} p(u_0^n, \mu_{S^c}, \chi_{S^c \setminus T}, y^n) \\
& \times \mathbf{1} \left\{ \exists \mathbf{z} \in \prod_{j \in B} [2^{nC_{jd}}] : (\forall j \in B : \mu_j(z_j) = u_j^n) \wedge (\forall j \in B \setminus T : \chi_j(z_j) = x_j^n) \right\} \\
& = \Pr \left(\{U_0^n = u_0^n, Y^n = y^n\} \cap \bigcup_{\mathbf{z}} E_{\mathbf{z}} \right) \\
& = \Pr \left(\bigcup_{\mathbf{z}} (\{U_0^n = u_0^n, Y^n = y^n\} \cap E_{\mathbf{z}}) \right) \\
& \stackrel{(a)}{\leq} 2^{nC_{Bd}} \Pr \left(\{U_0^n = u_0^n, Y^n = y^n\} \cap E_{\mathbf{z}=1} \right) \\
& = 2^{nC_{Bd}} \tilde{p}(u_0^n, u_B^n, x_{B \setminus T}^n, y^n) \\
& \stackrel{(b)}{=} 2^{nC_{Bd}} p(u_0^n) p_{\text{ind}}(u_B^n, x_{B \setminus T}^n | u_0^n) \tilde{p}(y^n | u_0^n, u_B^n, x_{B \setminus T}^n), \tag{26}
\end{aligned}$$

where (a) follows by the union bound and (b) follows by (25). Using a similar argument we can show

$$\sum_{\mu_A} p(\mu_A | u_0^n) \mathbf{1} \left\{ \exists \mathbf{z} \in \prod_{j \in A} [2^{nC_{jd}}] : \forall j \in A, \mu_j(z_j) = u_j^n \right\} \leq 2^{nC_{Ad}} p_{\text{ind}}(u_A^n | u_0^n). \tag{27}$$

Thus by (24), (26), and (27), the expression

$$\begin{aligned}
& 2^{n(\sum_{j \in A} R_{jd} + \sum_{j \in A \cup (B \cap T)} R_{jj} + C_{Ad} + C_{Bd})} \\
& \times \sum_{A_\epsilon^{(n)}} p(u_0^n) p_{\text{ind}}(u_{A \cup B}^n | u_0^n) p(x_{A \cup B}^n | u_0^n, u_{A \cup B}^n) \tilde{p}(y^n | u_0^n, u_B^n, x_{B \setminus T}^n)
\end{aligned}$$

is an upper bound for (23). Applying Lemma 11 to $p_{\text{ind}}(u_{A \cup B}^n | u_0^n)$ and dropping the epsilon term, this expression can be further bounded from above by

$$\begin{aligned}
& 2^{n(\sum_{j \in A} R_{jd} + \sum_{j \in A \cup (B \cap T)} R_{jj} + \zeta_{(A \cup B) \cap S_d})} \\
& \times \sum_{A_\epsilon^{(n)}(U_0, U_B, X_{B \setminus T}, Y)} p(u_0^n, u_B^n, x_{B \setminus T}^n) \tilde{p}(y^n | u_0^n, u_B^n, x_{B \setminus T}^n) \\
& \times \sum_{A_\epsilon^{(n)}(u_0^n, u_B^n, x_{B \setminus T}^n, y^n)} p(u_A^n | u_0^n, u_B^n) p(x_{A \cup (B \cap T)}^n | u_0^n, u_{A \cup (B \cap T)}^n)
\end{aligned}$$

Using (17), we can further upper bound the logarithm of this expression by

$$\begin{aligned}
& n \left[\sum_{j \in A} R_{jd} + \sum_{j \in A \cup (B \cap T)} R_{jj} + \zeta_{(A \cup B) \cap S_d} \right] \\
& + \log \sum_{A_\epsilon^{(n)}(U_0, U_B, X_{B \setminus T}, Y)} p(u_0^n, u_B^n, x_{B \setminus T}^n) \tilde{p}(y^n | u_0^n, u_B^n, x_{B \setminus T}^n) \\
& - nH(U_A | U_0, U_B) - nH(X_{A \cup (B \cap T)} | U_0, U_{A \cup (B \cap T)}) \\
& + nH(U_A, X_{A \cup (B \cap T)} | U_0, U_B, X_{B \setminus T}, Y)
\end{aligned}$$

Hence $\Pr(\mathcal{E}_{S,T}^{A,B}) \rightarrow 0$ if

$$\begin{aligned}
& \sum_{j \in A} R_{jd} + \sum_{j \in A \cup (B \cap T)} R_{jj} \\
& < -\zeta_{(A \cup B) \cap S_d} + H(U_A|U_0, U_B) + H(X_{A \cup (B \cap T)}|U_0, U_{A \cup (B \cap T)}) \\
& \quad - H(U_A, X_{A \cup (B \cap T)}|U_0, U_B, X_{B \setminus T}, Y) \\
& = I(U_A, X_{A \cup (B \cap T)}; Y|U_0, U_B, X_{B \setminus T}) - \zeta_{(A \cup B) \cap S_d},
\end{aligned}$$

where the last equality follows from the fact that

$$\begin{aligned}
H(U_A|U_0, U_B) &= H(U_A|U_0, U_B, X_{B \setminus T}) + I(U_A; X_{B \setminus T}|U_0, U_B) \\
&= H(U_A|U_0, U_B, X_{B \setminus T})
\end{aligned}$$

and

$$\begin{aligned}
H(X_{A \cup (B \cap T)}|U_0, U_{A \cup B}) &= H(X_{A \cup (B \cap T)}|U_0, U_{A \cup B}, X_{B \setminus T}) + I(X_{A \cup (B \cap T)}; X_{B \setminus T}|U_0, U_{A \cup B}) \\
&= H(X_{A \cup (B \cap T)}|U_0, U_{A \cup B}, X_{B \setminus T}).
\end{aligned}$$

Thus $\Pr(\mathcal{E}_{S,T}) \rightarrow 0$ if for some $S \cap S_d^c \subseteq A \subseteq S$ and $S^c \cap S_d^c \subseteq B \subseteq S^c$ such that $A \cup (B \cap T) \neq \emptyset$,

$$\begin{aligned}
& \sum_{j \in A} R_{jd} + \sum_{j \in A \cup (B \cap T)} R_{jj} \\
& < I(U_A, X_{A \cup (B \cap T)}; Y|U_0, U_B, X_{B \setminus T}) - \zeta_{(A \cup B) \cap S_d}
\end{aligned} \tag{28}$$

The bounds we obtain above are in terms of $(R_{jd})_{j=1}^k$ and $(R_{jj})_{j=1}^k$. To convert these to bounds in terms of $(R_j)_{j=1}^k$, recall that $R_{j0} = \min\{C_{j0}, R_j\}$, $R_{jj} = (R_j - C_{\text{in}}^j)^+$, and

$$\begin{aligned}
R_{jd} &= R_j - R_{j0} - R_{jj} \\
&= R_j - \min\{C_{j0}, R_j\} - R_{jj} = \max\{R_j - C_{j0}, 0\} - (R_j - C_{\text{in}}^j)^+ \\
&= (R_j - C_{j0})^+ - (R_j - C_{\text{in}}^j)^+.
\end{aligned}$$

Thus (28) can be written as

$$\begin{aligned}
& \sum_{j \in A} (R_j - C_{j0})^+ + \sum_{j \in B \cap T} (R_j - C_{\text{in}}^j)^+ \\
& < I(U_A, X_{A \cup (B \cap T)}; Y|U_0, U_B, X_{B \setminus T}) - \zeta_{(A \cup B) \cap S_d}
\end{aligned}$$

B. Theorem 3 (Sum-capacity gain)

Fix any unit vector $\mathbf{v} \in \mathbb{R}_{>0}^k$, rate vector $\mathbf{C}_{\text{in}} \in \mathbb{R}_{>0}^k$, and $\mathbf{B} \in \mathbb{R}_{\geq 0}^k$. For every $h \geq 0$, define $\mathbf{C}_{\text{out}}(h) = h\mathbf{v}$. In the achievable region defined in Section II, let $\mathcal{U}_0 = \{0, 1\}$, and for every $j \in [k]$, let $\mathcal{U}_j = \mathcal{X}_j$. Set $C_{j0} = 0$ and $C_{jd} = C_{\text{out}}^j(h)$ for every $j \in [k]$. For $h > 0$, let $\mathcal{P}(h)$ be the set of all distributions of the form

$$p(u_0, u_{[k]}) \cdot \prod_{j \in [k]} p(x_j|u_0, u_j)$$

that satisfy dependence constraints

$$\sum_{j \in S} C_{\text{out}}^j(h) - \sum_{j \in S} H(U_j|U_0) + H(U_S|U_0) > 0 \quad \forall \emptyset \subsetneq S \subseteq [k],$$

and cost constraints

$$\mathbb{E}[b_j(X_j)] \leq B_j \quad \forall j \in [k].$$

Using Lemma 13 (see end of this section), we see that every rate vector $(R_j)_{j \in [k]}$ that for some distribution $p \in \mathcal{P}(h)$ and every pair of subsets $S, T \subseteq [k]$ satisfies

$$\sum_{j \in S \cup T} R_j < I(X_{S \cup T}; Y | U_0, U_{S^c}, X_{S^c \cap T^c}) + \sum_{j \in T \setminus S} C_{\text{in}}^j - \zeta_{[k]} \quad (29)$$

and

$$\sum_{j \in [k]} R_j < I(X_{[k]}; Y) - \zeta_{[k]},$$

is achievable. This follows from setting $A = S$ and $B = S^c$ for every $S, T \subseteq [k]$ in (8). To obtain a lower bound on sum-capacity, we evaluate this region for a specific distribution in $\mathcal{P}(h)$.

Since our MAC is in $\mathcal{C}^*(\mathcal{X}_{[k]}, \mathcal{Y})$, there exists a distribution $p_a \in \mathcal{P}_{\text{ind}}(\mathcal{X}_{[k]})$ that satisfies

$$I_a(X_{[k]}; Y) = \max_{p \in \mathcal{P}_{\text{ind}}(\mathcal{X}_{[k]})} I(X_{[k]}; Y),$$

and a distribution $p_b \in \mathcal{P}(\mathcal{X}_{[k]})$ that satisfies

$$\mathbb{E}_b \left[D(p(y|X_{[k]}) \| p_a(y)) \right] > \mathbb{E}_a \left[D(p(y|X_{[k]}) \| p_a(y)) \right],$$

and whose support is contained in the support of p_a . Here we also assume that for all $j \in [k]$,

$$I_a(X_j; Y | X_{[k] \setminus \{j\}}) > 0.$$

At the end of the proof, we show that in the case where this property does not hold, the same result follows by considering a MAC with a smaller number of users.

Choose $\mu \in (0, 1)$ such that for every nonempty $S \subseteq [k]$,

$$\mu I_a(X_S; Y | X_{S^c}) < \sum_{j \in S} C_{\text{in}}^j. \quad (30)$$

For every $\lambda \in [0, 1]$, define the distribution $p_\lambda(u_0, u_{[k]}, x_{[k]})$ as

$$p_\lambda(u_0, u_{[k]}, x_{[k]}) = p_\lambda(u_0) p_\lambda(u_{[k]}) p_\lambda(x_{[k]} | u_0, u_{[k]}),$$

where

$$p_\lambda(u_0) = \begin{cases} \mu & \text{if } u_0 = 1 \\ 1 - \mu & \text{if } u_0 = 0, \end{cases}$$

and for every $u_{[k]} \in \mathcal{U}_{[k]}$ (recall $\mathcal{U}_{[k]} = \mathcal{X}_{[k]}$),

$$p_\lambda(u_{[k]}) = (1 - \lambda) p_a(u_{[k]}) + \lambda p_b(u_{[k]}).$$

Finally, for every $(u_0, u_{[k]}, x_{[k]})$,

$$p_\lambda(x_{[k]}|u_0, u_{[k]}) = \prod_{j=1}^k p_\lambda(x_j|u_0, u_j),$$

where for all $j \in [k]$,

$$p_\lambda(x_j|u_0, u_j) = \begin{cases} \mathbf{1}\{x_j = u_j\} & \text{if } u_0 = 1 \\ p_a(x_j) & \text{if } u_0 = 0. \end{cases}$$

Note that $p_\lambda(u_0)$ and $p_\lambda(x_{[k]}|u_0, u_{[k]})$ do not depend on λ . In addition, since p_a and p_b satisfy the cost constraints and

$$p_\lambda(x_{[k]}) = (1 - \lambda)p_a(x_{[k]}) + \lambda p_b(x_{[k]}),$$

for all $\lambda \in (0, 1)$, p_λ satisfies the cost constraints as well.

We next find a function $\lambda^*(h)$ so that

$$p_{\lambda^*(h)}(u_0, u_{[k]}, x_{[k]}) \in \mathcal{P}(h)$$

for sufficiently small h . Fix $\epsilon > 0$, and consider the equation

$$h \sum_{j \in [k]} v_j = \sum_{j \in [k]} H_\lambda(U_j) - H_\lambda(U_{[k]}) + \epsilon \lambda \sum_{j \in [k]} v_j. \quad (31)$$

By Lemma 14 (see end of this section),

$$\left. \frac{dh}{d\lambda} \right|_{\lambda=0^+} = \epsilon > 0.$$

Thus the inverse function theorem implies that there exists a function $\lambda = \lambda^*(h)$ defined on $[0, h_0)$ for some $h_0 > 0$ that satisfies (31), and

$$\left. \frac{d\lambda^*}{dh} \right|_{h=0^+} = \frac{1}{\epsilon}. \quad (32)$$

For every nonempty $S \subseteq [k]$, define the function $\zeta_S : [0, h_0) \rightarrow \mathbb{R}$ as

$$\zeta_S(h) = \sum_{j \in S} C_{\text{out}}^j(h) - \sum_{j \in S} H_{\lambda^*}(U_j) + H_{\lambda^*}(U_S), \quad (33)$$

If we calculate the derivative of ζ_S at $h = 0$, by Lemma 14, we get

$$\left. \frac{d\zeta_S}{dh} \right|_{h=0^+} = \sum_{j \in S} v_j > 0.$$

This implies that there exists $0 < h_1 \leq h_0$ such that for every $0 < h < h_1$ and all nonempty $S \subseteq [k]$,

$$\zeta_S(h) > 0.$$

Therefore, for all sufficiently small h , $p_{\lambda^*(h)}(u_0, u_{[k]}, x_{[k]})$ is in $\mathcal{P}(h)$.

We next find a lower bound for the achievable sum-rate using the distribution $p_{\lambda^*}(u_0, u_{[k]}, x_{[k]})$ for small h . For every $S, T \subseteq [k]$, define the function $f_{S,T} : [0, h_1) \rightarrow \mathbb{R}$ as

$$f_{S,T}(h) = I_{\lambda^*}(X_{S \cup T}; Y|U_0, U_{S^c}, X_{S^c \cap T^c}) + \sum_{j \in T \setminus S} C_{\text{in}}^j - \zeta_{[k]}(h).$$

In the above equation, expanding the mutual information term with respect to U_0 gives

$$\begin{aligned} I_{\lambda^*}(X_{S \cup T}; Y | U_0, U_{S^c}, X_{S^c \cap T^c}) \\ = \mu I_{\lambda^*}(X_S; Y | X_{S^c}) + (1 - \mu) I_a(X_{S \cup T}; Y | X_{S^c \cap T^c}), \end{aligned}$$

where the term $I_{\lambda^*}(X_S; Y | X_{S^c})$ is calculated with respect to the distribution

$$(1 - \lambda)p_a(x_{[k]}) + \lambda p_b(x_{[k]}).$$

Next, for every $S \subseteq [k]$, define the function $F_S : [0, h_1] \rightarrow \mathbb{R}$ as

$$F_S(h) = I_{\lambda^*}(X_S; Y | U_0, U_{S^c}, X_{S^c}) - \zeta_{[k]}(h).$$

The following argument shows that for sufficiently small h and for all $S, T \subseteq [k]$,

$$f_{S,T}(h) \geq F_{S \cup T}(h).$$

Consider some S and T for which $T \setminus S$ is not empty. Then

$$\begin{aligned} f_{S,T}(0) &= \mu I_a(X_S; Y | X_{S^c}) + (1 - \mu) I_a(X_{S \cup T}; Y | X_{S^c \cap T^c}) + \sum_{j \in T \setminus S} C_{\text{in}}^j \\ &\stackrel{(*)}{>} \mu I_a(X_S; Y | X_{S^c}) + (1 - \mu) I_a(X_{S \cup T}; Y | X_{S^c \cap T^c}) + \mu I_a(X_{T \setminus S}; Y | X_{(T \setminus S)^c}) \\ &\geq I_a(X_{S \cup T}; Y | X_{S^c \cap T^c}) = F_{S \cup T}(0), \end{aligned}$$

where $(*)$ follows from (30). Note that $f_{S,T}$ and $F_{S \cup T}$ are continuous functions of h for all S and T . Thus there exists $0 < h_2 \leq h_1$ such that for every $h \in [0, h_2)$ and $S, T \subseteq [k]$ with $T \setminus S \neq \emptyset$,

$$f_{S,T}(h) \geq F_{S \cup T}(h).$$

Next consider S and T for which $T \setminus S$ is empty; that is, T is a subset of S . In this case

$$\begin{aligned} f_{S,T}(h) &= I_{\lambda^*}(X_{S \cup T}; Y | U_0, U_{S^c}, X_{S^c \cap T^c}) + \sum_{j \in T \setminus S} C_{\text{in}}^j - \zeta_{[k]}(h) \\ &= I_{\lambda^*}(X_S; Y | U_0, U_{S^c}, X_{S^c}) - \zeta_{[k]}(h) \\ &= F_S(h) = F_{S \cup T}(h). \end{aligned}$$

Thus $f_{S,T}(h) \geq F_{S \cup T}(h)$ for all such S and T as well. Now fix $h \in [0, h_2)$. From the above argument, it follows that the set of all rate vectors that satisfy

$$0 \leq \sum_{j \in S} R_j \leq F_S(h) \quad \forall \emptyset \neq S \subseteq [k]$$

is achievable. Denote this region with $\mathcal{C}_{\text{ach}}(h)$. Now consider the set of all rate vectors that satisfy

$$0 \leq \sum_{j \in S} R_j \leq \Phi_S(h) \quad \forall \emptyset \neq S \subseteq [k],$$

where $\Phi_S(h)$ is defined as

$$\Phi_S(h) = F_S(h) + \zeta_{S^c}(h) + \sum_{j \in S} C_{\text{out}}^j(h).$$

Denote this set with $\mathcal{C}_{\text{out}}(h)$. Note that $\mathcal{C}_{\text{out}}(h)$ is an outer bound for $\mathcal{C}_{\text{ach}}(h)$.

We next show that there exists $0 < h_3 \leq h_2$ such that for every $j \in [k]$ and all $0 < h < h_3$,

$$\Phi_{\{j\}}(h) > k \sum_{i=1}^k C_{\text{out}}^i(h). \quad (34)$$

To see this, first note that the right hand side of the above equation equals zero at $h = 0$, while

$$\Phi_{\{j\}}(0) = I_a(X_j; Y | X_{[k] \setminus \{j\}}) > 0.$$

Inequality (34) now follows from the fact that both sides are continuous in h .

By Lemma 15, for a fixed h , the mapping $S \mapsto \Phi_S(h)$ is submodular and nondecreasing. Thus for every $j \in [k]$, there exists a rate vector $(R_i)_{i \in [k]}$ in $\mathcal{C}_{\text{out}}(h)$ such that

$$R_j > k \sum_{i=1}^k C_{\text{out}}^i(h),$$

and

$$\sum_{j \in [k]} R_j = \Phi_{[k]}(h).$$

For example, for $j = 1$, consider the rate vector $(R_i)_{i \in [k]}$, where $R_1 = \Phi_{\{1\}}(h)$, and for all $i > 1$,

$$R_i = \Phi_{[i]} - \Phi_{[i-1]}.$$

From Corollary 44.3a in [20, pp. 772] it follows that the defined rate vector is in $\mathcal{C}_{\text{out}}(h)$. Now since $\mathcal{C}_{\text{out}}(h)$ is a convex region, it follows that there exists a rate vector $(R_j^*(h))_j$ such that for all $j \in [k]$,

$$R_j^*(h) > \sum_{j=1}^k C_{\text{out}}^j(h),$$

and

$$\sum_{j=1}^k R_j^*(h) = \Phi_{[k]}(h).$$

On the other hand, from the definition of $\zeta_S(h)$, given by (33), it follows

$$\Phi_S(h) \leq F_S(h) + \sum_{j=1}^k C_{\text{out}}^j(h).$$

Thus

$$\left(R_j^*(h) - \sum_{j=1}^k C_{\text{out}}^j(h) \right)_{j \in [k]} \in \mathcal{C}_{\text{ach}}(h).$$

This implies that the sum-rate

$$\begin{aligned} R_{\text{sum}}(h) &= \Phi_{[k]}(h) - k \sum_{j=1}^k C_{\text{out}}^j(h) \\ &= \mu I_{\lambda^*}(X_{[k]}; Y) + (1 - \mu) I_a(X_{[k]}; Y) - k \sum_{j=1}^k C_{\text{out}}^j(h) \end{aligned}$$

is achievable. In addition, since

$$R_{\text{sum}}(0) = I_a(X_{[k]}; Y) = \max_{p \in \mathcal{P}_{\text{ind}}(\mathcal{X}_{[k]})} I(X_{[k]}; Y),$$

we have

$$G(h\mathbf{v}) \geq R_{\text{sum}}(h) - R_{\text{sum}}(0) \quad (35)$$

for all $h \in [0, h_3)$. Thus

$$\begin{aligned} (D_{\mathbf{v}}G)(\mathbf{0}) &= \lim_{h \rightarrow 0^+} \frac{G(h\mathbf{v})}{h} \\ &\stackrel{(i)}{\geq} \lim_{h \rightarrow 0^+} \frac{R_{\text{sum}}(h) - R_{\text{sum}}(0)}{h} \\ &= \mu \frac{d}{d\lambda^*} I_{\lambda^*}(X_{[k]}; Y) \Big|_{\lambda^*=0^+} \times \frac{d\lambda^*}{dh} \Big|_{h=0^+} - k \sum_{j=1}^k v_j \\ &\stackrel{(ii)}{\geq} \frac{\mu}{\epsilon} \left[\sum_{x_{[k]}} (p_b(x_{[k]}) - p_a(x_{[k]})) D(p(y|x_{[k]}) \| p_a(y)) \right] - k \sum_{j=1}^k v_j. \end{aligned} \quad (36)$$

Here (i) follows from (35) and (ii) is proved by combining (32) and Lemma 14, which appears at the end of this section. From our definitions of p_a and p_b it follows

$$\sum_{x_{[k]}} p_b(x_{[k]}) D(p(y|x_{[k]}) \| p_a(y)) > \sum_{x_{[k]}} p_a(x_{[k]}) D(p(y|x_{[k]}) \| p_a(y)).$$

Since ϵ is arbitrary, from (36) we get

$$(D_{\mathbf{v}}G)(\mathbf{0}) = \infty.$$

This completes the proof for the case where

$$S_* := \{j \in [k] : I_a(X_j; Y | X_{[k] \setminus \{j\}}) > 0\}$$

contains $[k]$ (i.e., $S_* = [k]$). We next consider a MAC for which S_* is a strict subset of $[k]$ (i.e., $S_* \subsetneq [k]$).

For every $j \in [k]$, let $\mathcal{A}_j \subseteq \mathcal{X}_j$ denote the support of $p_a(x_j)$. Then for nonempty $S \subseteq [k]$, the support of $p_a(x_S)$ is given by

$$\mathcal{A}_S = \prod_{j \in S} \mathcal{A}_j.$$

Note that

$$I_a(X_{S_*^c}; Y | X_{S_*}) \leq \sum_{j \in S_*^c} I_a(X_j; Y | X_{[k] \setminus \{j\}}) = 0.$$

Thus for every $x_{S_*} \in \mathcal{A}_{S_*}$,

$$I_a(X_{S_*^c}; Y | X_{S_*} = x_{S_*}) = 0,$$

which implies for all $x_{[k]} \in \mathcal{A}_{[k]}$,

$$p(y|x_{[k]}) = p_a(y|x_{S_*}).$$

Note that since the support of p_b is contained in the support of p_a by assumption, it follows that for all nonempty $S \subseteq [k]$, the support of $p_b(x_S)$ is contained in \mathcal{A}_S .

Now consider the $|S_*|$ -user MAC

$$(\mathcal{A}_{S_*}, p_a(y|x_{S_*}), \mathcal{Y}),$$

and the input distributions $p_{\text{ind}}(x_{S_*}) = p_a(x_{S_*})$ and $p_{\text{dep}}(x_{S_*}) = p_b(x_{S_*})$. Note that

$$I_{\text{ind}}(X_{S_*}; Y) = \max_{p \in \mathcal{P}(\mathcal{X}_{S_*})} I(X_{S_*}; Y),$$

and

$$\begin{aligned} \mathbb{E}_{\text{dep}} \left[D(p_a(y|X_{S_*}) \| p_{\text{ind}}(y)) \right] &= \mathbb{E}_b \left[D(p(y|X_{[k]}) \| p_a(y)) \right] \\ &> \mathbb{E}_a \left[D(p(y|X_{[k]}) \| p_a(y)) \right] \\ &= \mathbb{E}_{\text{ind}} \left[D(p_a(y|X_{S_*}) \| p_{\text{ind}}(y)) \right]. \end{aligned}$$

Furthermore, for every $j \in S_*$,

$$I_{\text{ind}}(X_j; Y | X_{S_* \setminus \{j\}}) = I_a(X_j; Y | X_{[k] \setminus \{j\}}) > 0.$$

Thus this MAC satisfies all of the conditions under which we already proved Theorem 3. Suppose $\mathbf{v} = (v_j)_{j=1}^k$ is a unit vector in $\mathbb{R}_{>0}^k$. Let

$$|\mathbf{v}_{S_*}| = \left(\sum_{j \in S_*} v_j^2 \right)^{1/2},$$

and define $\mathbf{v}^* = (v_j^*)_{j=1}^k \in \mathbb{R}_{>0}^k$ as

$$v_j^* = \frac{v_j}{|\mathbf{v}_{S_*}|} \mathbf{1}\{j \in S_*\}.$$

Then

$$\begin{aligned} (D_{\mathbf{v}}G)(\mathbf{0}) &= \lim_{h \rightarrow 0^+} \frac{G(h\mathbf{v})}{h} \\ &\geq |\mathbf{v}_{S_*}| \times \lim_{h \rightarrow 0^+} \frac{G(h|\mathbf{v}_{S_*}|\mathbf{v}^*)}{h|\mathbf{v}_{S_*}|} \stackrel{(\star)}{=} \infty, \end{aligned}$$

where (\star) follows from the fact that our $|S_*|$ -user MAC satisfies all the required properties to imply an infinite directional derivative for sum-capacity.

We next provide the proofs for the lemmas we use in the above argument.

The first lemma allows us to simplify the achievable region by replacing the terms $(R_j - C_{\text{in}}^j)^+$ with $R_j - C_{\text{in}}^j$.

Lemma 13. *Let k be a positive integer. Fix $\gamma > 0$ and for every $j \in [k]$, let α_j be a real number. Then the vector $(x_j)_{j \in [k]}$ satisfies*

$$\sum_{j \in [k]} (x_j - \alpha_j)^+ < \gamma$$

if and only if for every nonempty $S \subseteq [k]$,

$$\sum_{j \in S} (x_j - \alpha_j) < \gamma.$$

Proof: Define the sets \mathcal{A}^+ and \mathcal{A} as follows

$$\begin{aligned} \mathcal{A} &= \left\{ \mathbf{x} \mid \forall S \subseteq [k] : \sum_{j \in S} (x_j - \alpha_j) < \gamma \right\} \\ \mathcal{A}^+ &= \left\{ \mathbf{x} \mid \sum_{j \in [k]} (x_j - \alpha_j)^+ < \gamma \right\} \end{aligned}$$

Our aim is to show $\mathcal{A} = \mathcal{A}^+$. We first prove $\mathcal{A} \supseteq \mathcal{A}^+$. For every $j \in [k]$, $x_j - \alpha_j \leq (x_j - \alpha_j)^+$; thus $\mathcal{A}^+ \subseteq \mathcal{A}$. We next prove $\mathcal{A} \subseteq \mathcal{A}^+$. Consider any $\mathbf{x} \in \mathcal{A}$. Define the set $S \subseteq [k]$ as

$$S = \{j \in [k] | x_j > \alpha_j\}.$$

If $S = \emptyset$, then $\mathbf{x} \in \mathcal{A}^+$ as $\gamma > 0$. If S is not empty, then

$$\sum_{j \in [k]} (x_j - \alpha_j)^+ = \sum_{j \in S} (x_j - \alpha_j) < \gamma.$$

Thus $\mathbf{x} \in \mathcal{A}^+$. ■

The next lemma provides the derivative of the input-output mutual information and the total correlation [21], when calculated with respect to the convex combination of two distributions. In this lemma, $\mathcal{X}_{[k]}$ may be finite, countably infinite, or equal to \mathbb{R}^k . In the first two cases, p_a and p_b are probability mass functions. In the case where $\mathcal{X}_{[k]} = \mathbb{R}^k$, we assume p_a and p_b are “bounded” probability density functions. We say a probability density function $p(x_{[k]})$ on \mathbb{R}^k is bounded if

$$\forall \emptyset \subsetneq S \subseteq [k] : \sup_{\mathcal{X}_S} p(x_S) < \infty.$$

In addition, in the case where $\mathcal{X}_{[k]} = \mathbb{R}^k$, the sums should be replaced with integrals.

Lemma 14. *Consider two distributions p_a and p_b defined on $\mathcal{X}_{[k]}$. For every $\lambda \in [0, 1]$, define the distribution p_λ on $\mathcal{X}_{[k]}$ as*

$$p_\lambda(x_{[k]}) = (1 - \lambda)p_a(x_{[k]}) + \lambda p_b(x_{[k]}).$$

Then the following statements are true.

(i) *For every nonempty $S \subseteq [k]$, we have*

$$\frac{d}{d\lambda} H_\lambda(X_S) = - \sum_{x_S} (p_b(x_S) - p_a(x_S)) \log p_\lambda(x_S).$$

(ii) *For every k -user MAC $(\mathcal{X}_{[k]}, p(y|x_{[k]}), \mathcal{Y})$, we have*

$$\frac{d}{d\lambda} I_\lambda(X_{[k]}; Y) = \sum_{x_{[k]}} (p_b(x_{[k]}) - p_a(x_{[k]})) D(p(y|x_{[k]}) || p_\lambda(y)). \quad (37)$$

(iii) *If p_a has the form*

$$p_a(x_{[k]}) = \prod_{j \in [k]} p_a(x_j),$$

and the support of $p_a(x_{[k]})$ contains the support of $p_b(x_{[k]})$, then for every nonempty $S \subseteq [k]$,

$$\frac{d}{d\lambda} \left(\sum_{j \in S} H_\lambda(X_j) - H_\lambda(X_S) \right) \Big|_{\lambda=0^+} = 0. \quad (38)$$

Proof:

Claim (i) is clear in the case where \mathcal{X}_S is finite. In the case where \mathcal{X}_S is infinite, we apply the dominated convergence theorem [22, p. 55]. Define $f : \mathcal{X}_S \times [0, 1] \rightarrow \mathbb{R}$ as

$$f(x_S, \lambda) = p_\lambda(x_S) \log \frac{1}{p_\lambda(x_S)}.$$

Fix $\lambda \in [0, 1]$, and consider the sequence of functions $g_n(x_S)$ defined as

$$g_n(x_S) = n \left(f(x_S, \lambda + \frac{1}{n}) - f(x_S, \lambda) \right).$$

For all $x_S \in \mathcal{X}_S$, we have

$$\lim_{n \rightarrow \infty} g_n(x_S) = \frac{\partial f}{\partial \lambda}(x_S, \lambda) = -(\log e + p_\lambda(x_S))(p_b(x_S) - p_a(x_S)).$$

By the mean value theorem, for all $x_S \in \mathcal{X}_S$ and $n \in \mathbb{Z}_{>0}$, there exists $h' \in (0, 1/n)$ such that

$$g_n(x_S) = \frac{\partial f}{\partial \lambda}(x_S, \lambda + h') = -(\log e + p_{\lambda+h'}(x_S))(p_b(x_S) - p_a(x_S)).$$

Since p_a and p_b are bounded, so is $p_{\lambda+h'}$, and thus, for some constant $C > 0$ and all $n \in \mathbb{Z}_{>0}$,

$$|g_n(x_S)| \leq C |p_b(x_S) - p_a(x_S)|.$$

Define $\varphi : \mathcal{X}_S \rightarrow \mathbb{R}$ as

$$\varphi(x_S) = C |p_b(x_S) - p_a(x_S)|.$$

Note that $\varphi \in L^1(\mathcal{X}_S)$, since

$$\int_{\mathcal{X}_S} |\varphi(x_S)| dx_S \leq 2C.$$

By the dominated convergence theorem,

$$\lim_{n \rightarrow \infty} \sum_{x_S \in \mathcal{X}_S} g_n(x_S) = \sum_{x_S \in \mathcal{X}_S} \lim_{n \rightarrow \infty} g_n(x_S),$$

which implies

$$\frac{d}{d\lambda} H_\lambda(X_S) = - \sum_{x_S} (p_b(x_S) - p_a(x_S)) \log p_\lambda(x_S).$$

For (ii), note that

$$p_\lambda(y) = (1 - \lambda)p_a(y) + \lambda p_b(y).$$

Thus by (i),

$$\begin{aligned} \frac{d}{d\lambda} H_\lambda(Y) &= - \sum_y (p_b(y) - p_a(y)) (\log e + \log p_\lambda(y)) \\ &= \sum_y (p_b(y) - p_a(y)) \log \frac{1}{p_\lambda(y)} \\ &= \sum_{x_{[k]}} (p_b(x_{[k]}) - p_a(x_{[k]})) \sum_y p(y|x_{[k]}) \log \frac{1}{p_\lambda(y)}. \end{aligned}$$

Similarly,

$$\begin{aligned} \frac{d}{d\lambda} H_\lambda(Y|X_{[k]}) &= \sum_{x_{[k]}} (p_b(x_{[k]}) - p_a(x_{[k]})) \sum_y p(y|x_{[k]}) \log \frac{1}{p(y|x_{[k]})}. \end{aligned}$$

Taking the difference between these derivatives completes the proof of part (ii).

For part (iii), note that for every $j \in [k]$,

$$\begin{aligned} \frac{d}{d\lambda} H_\lambda(X_j) &= - \sum_{x_j} (p_b(x_j) - p_a(x_j)) (\log e + \log p_\lambda(x_j)) \\ &= - \sum_{x_j} (p_b(x_j) - p_a(x_j)) \log p_\lambda(x_j) \end{aligned}$$

and

$$\begin{aligned} \frac{d}{d\lambda} \sum_{j \in S} H_\lambda(X_j) &= - \sum_{j \in S} \sum_{x_j} (p_b(x_j) - p_a(x_j)) \log p_\lambda(x_j) \\ &= \sum_{x_S} (p_b(x_S) - p_a(x_S)) \log \frac{1}{\prod_{j \in S} p_\lambda(x_j)} \end{aligned}$$

On the other hand,

$$\frac{d}{d\lambda} H_\lambda(X_S) = - \sum_{x_S} (p_b(x_S) - p_a(x_S)) \log p_\lambda(x_S).$$

Thus

$$\frac{d}{d\lambda} \left(\sum_{j \in S} H_\lambda(X_j) - H_\lambda(X_S) \right) = \sum_{x_S} (p_b(x_S) - p_a(x_S)) \log \frac{p_\lambda(x_S)}{\prod_{j \in S} p_\lambda(x_j)}.$$

Equation (38) now follows from the fact that

$$p_a(x_S) = \prod_{j \in S} p_a(x_j),$$

and the support of p_b is contained in the support of p_a . ■

In the next lemma, we prove that for a fixed h , the mapping $S \mapsto \Phi_S(h)$ is nondecreasing and submodular. In the statement of this lemma, $2^{[k]}$ denotes the collection of all subsets of $[k]$.

Lemma 15. *Fix a distribution*

$$p(u_{[k]}) \cdot \prod_{j=1}^k p(x_j | u_j) \cdot p(y | x_{[k]})$$

on $\mathcal{U}_{[k]} \times \mathcal{X}_{[k]} \times \mathcal{Y}$, and define the function $\Phi : 2^{[k]} \rightarrow \mathbb{R}$ as

$$\Phi(S) = I(X_S; Y | U_{S^c} X_{S^c}) + \sum_{j \in S} H(U_j) - H(U_S | U_{S^c})$$

for every $S \subseteq [k]$. Then Φ is nondecreasing and submodular.

Proof: Note that

$$\Phi(S) = H(Y | U_{S^c} X_{S^c}) - H(Y | X_{[k]}) + \sum_{j \in S} H(U_j) + H(U_{S^c}) - H(U_{[k]}).$$

For every $j \in [k]$, let $V_j = (U_j, X_j)$. Then for every $S \subseteq [k]$,

$$\begin{aligned} \sum_{j \in S} H(V_j) + H(V_{S^c}) - H(V_{[k]}) &= \sum_{j \in S} H(U_j, X_j) + H(U_{S^c}, X_{S^c}) - H(U_{[k]}, X_{[k]}) \\ &= \sum_{j \in S} H(U_j) + H(U_{S^c}) - H(U_{[k]}), \end{aligned}$$

since each X_j only depends on U_j . Thus

$$\begin{aligned} \Phi(S) &= H(Y|V_{S^c}) - H(Y|V_{[k]}) + \sum_{j \in S} H(V_j) + H(V_{S^c}) - H(V_{[k]}) \\ &= H(V_{S^c}|Y) + \sum_{j \in S} H(V_j) - H(V_{[k]}|Y). \end{aligned}$$

We first show Φ is nondecreasing. Let S be a subset of T . Then

$$\begin{aligned} H(V_{S^c}|Y) + \sum_{j \in S} H(V_j) &= H(V_{T^c}|Y) + H(V_{S^c \setminus T^c}|V_{T^c}, Y) + \sum_{j \in T} H(V_j) - \sum_{j \in T \setminus S} H(V_j) \\ &\leq H(V_{T^c}|Y) + \sum_{j \in T} H(V_j), \end{aligned}$$

since

$$H(V_{S^c \setminus T^c}|V_{T^c}, Y) = H(V_{T \setminus S}|V_{T^c}, Y) \leq \sum_{j \in T \setminus S} H(V_j).$$

Thus Φ is nondecreasing.

We next show Φ is submodular. Fix $S, T \subseteq [k]$. Our aim is to prove

$$\Phi(S) + \Phi(T) \geq \Phi(S \cup T) + \Phi(S \cap T). \quad (39)$$

We have

$$\begin{aligned} H(V_{S^c}|Y) + H(V_{T^c}|Y) &= H(V_{S^c \cap T^c}|Y) + H(V_{S^c \setminus T^c}|V_{S^c \cap T^c}, Y) \\ &\quad + H(V_{S^c \cup T^c}|Y) - H(V_{S^c \setminus T^c}|V_{T^c}, Y) \\ &= H(V_{S^c \cap T^c}|Y) + H(V_{S^c \cup T^c}|Y) + I(V_{S^c \setminus T^c}; V_{T^c \setminus S^c} | V_{S^c \cap T^c}, Y) \\ &\geq H(V_{S^c \cap T^c}|Y) + H(V_{S^c \cup T^c}|Y). \end{aligned}$$

This proves (39), since

$$\sum_{j \in S} H(V_j) + \sum_{j \in T} H(V_j) = \sum_{j \in S \cup T} H(V_j) + \sum_{j \in S \cap T} H(V_j).$$

■

C. Proposition 4 (The k -user Gaussian MAC)

For the k -user Gaussian MAC, define p_{ind} as

$$p_{\text{ind}}(x_{[k]}) = \prod_{j \in [k]} \frac{1}{\sqrt{2\pi P_j}} \exp\left(-\frac{x_j^2}{2P_j}\right)$$

Note that p_{ind} satisfies

$$I_{\text{ind}}(X_{[k]}; Y) = \max_{p \in \mathcal{P}_{\text{ind}}(\mathcal{X}_{[k]})} I(X_{[k]}; Y).$$

From [23, p. 33],

$$\begin{aligned} & D(p(y|x_{[k]})||p_{\text{ind}}(y)) \\ &= \frac{1}{2} \left[\frac{1}{\sum_{j \in [k]} P_j + N} \left(\sum_{j \in [k]} x_j \right)^2 - \frac{\sum_{j \in [k]} P_j}{\sum_{j \in [k]} P_j + N} + \log \left(1 + \frac{1}{N} \sum_{j \in [k]} P_j \right) \right]. \end{aligned}$$

For p_{dep} , choose any density function that satisfies

$$\forall j \in [k] : \mathbb{E}_{\text{dep}}[|X_j|^2] \leq P_j$$

and

$$\mathbb{E}_{\text{dep}} \left[\left(\sum_{j \in [k]} X_j \right)^2 \right] > \sum_{j \in [k]} P_j. \quad (40)$$

Then (40) guarantees

$$\mathbb{E}_{\text{dep}} \left[D(p(y|X_{[k]})||p_{\text{ind}}(y)) \right] > \mathbb{E}_{\text{ind}} \left[D(p(y|X_{[k]})||p_{\text{ind}}(y)) \right].$$

For example, we may choose $p_{\text{dep}}(x_{[k]})$ to be the distribution $\mathcal{N}(\mathbf{0}, \Sigma)$, where $\Sigma = (\Sigma_{ij})_{i,j \in [k]}$ is given by

$$\Sigma_{ij} = \begin{cases} \rho \sqrt{P_i P_j} & \text{if } i \neq j \\ P_i & \text{if } i = j, \end{cases}$$

where ρ is any number in $(0, 1]$.

D. Proposition 5 (Outer bound)

Consider a $((2^{nR_1}, \dots, 2^{nR_k}), n, L)$ -code for the MAC with a $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$ -CF. For every message vector $w_{[k]} = (w_1, \dots, w_k)$, $j \in [k]$, and $\ell \in [L]$, define

$$\begin{aligned} u_{j\ell} &= \varphi_{j\ell}(w_j, v_j^{\ell-1}) \\ v_{j\ell} &= \psi_{j\ell}(u_1^\ell, \dots, u_k^\ell), \end{aligned}$$

where $u_j^\ell = (u_{j1}, \dots, u_{j\ell})$ and $v_j^\ell = (v_{j1}, \dots, v_{j\ell})$, respectively. Also, for every nonempty $S \subseteq [k]$ and $\ell \in [L]$, let $u_{S\ell} = (u_{j\ell})_{j \in S}$ and $u_S^\ell = (u_j^\ell)_{j \in S}$. Finally, for every $j \in [k]$, $\ell \in [L]$, and $v_j^{\ell-1} \in \mathcal{V}_j^{\ell-1}$, define the mapping

$$\begin{aligned} \varphi_{j\ell, v_j^{\ell-1}}^{-1} : \mathcal{U}_{j\ell} &\rightarrow 2^{[2^{nR_j}]} \\ u_{j\ell} &\mapsto \{w_j \mid \varphi_{j\ell}(w_j, v_j^{\ell-1}) = u_{j\ell}\}, \end{aligned}$$

where $2^{[2^{nR_j}]}$ denotes the set of all the subsets of $[2^{nR_j}]$.

Note that $(v_j^L)_{j=1}^k$ is a deterministic function of $u_{[k]}^L$. Thus for every $u_{[k]}^L$ and $j \in [k]$, the set

$$\mathcal{A}_j(u_{[k]}^L) = \bigcap_{\ell=1}^L \varphi_{j\ell, v_j^{\ell-1}}^{-1}(u_{j\ell})$$

is well-defined. It follows that for a fixed code and a given message vector $w_{[k]}$, the vector of all CF inputs is given by $u_{[k]}^L$ if and only if for every $j \in [k]$, $w_j \in \mathcal{A}_j(u_{[k]}^L)$.

By Fano's inequality [13, p. 38], for some $\epsilon_n = o(1)$,

$$H(W_{[k]}|Y^n) \leq n\epsilon_n.$$

Thus for every nonempty subset $S \subseteq [k]$,

$$H(W_S|W_{S^c}, U_{[k]}^L, Y^n) \leq n\epsilon_n.$$

We have

$$\begin{aligned} n \sum_{j \in S} R_j &\leq H(W_S|W_{S^c}) \\ &= I(W_S; U_{[k]}^L, Y^n|W_{S^c}) + H(W_S|W_{S^c}, U_{[k]}^L, Y^n) \\ &\leq I(W_S; U_{[k]}^L|W_{S^c}) + I(W_S; Y^n|W_{S^c}, U_{[k]}^L) + n\epsilon_n. \end{aligned} \quad (41)$$

We next find an upper bound for each of the mutual information terms. For the first term, we have

$$\begin{aligned} I(W_S; U_{[k]}^L|W_{S^c}) &\stackrel{(a)}{=} H(U_{[k]}^L|W_{S^c}) \\ &= \sum_{\ell=1}^L H(U_{[k]\ell}|W_{S^c}, U_{[k]}^{\ell-1}) \\ &= \sum_{\ell=1}^L H(U_{S\ell}, U_{S^c\ell}|W_{S^c}, U_{[k]}^{\ell-1}) \\ &\stackrel{(b)}{=} \sum_{\ell=1}^L H(U_{S\ell}|W_{S^c}, U_{[k]}^{\ell-1}, U_{S^c\ell}) \leq \sum_{j \in S} C_{\text{in}}^j, \end{aligned}$$

where (a) follows from the fact that $U_{[k]}^L$ is a deterministic function of $W_{[k]}$. Statement (b) follows from the fact that $U_{S^c\ell}$ is a deterministic function of $(W_{S^c}, U_{[k]}^{\ell-1})$. For the second term in (41), we have

$$\begin{aligned} I(W_S; Y^n|W_{S^c}, U_{[k]}^L) &= H(Y^n|W_{S^c}, U_{[k]}^L) - H(Y^n|W_S, W_{S^c}, U_{[k]}^L) \\ &= H(Y^n|U_{[k]}^L, X_{S^c}^n) - H(Y^n|U_{[k]}^L, X_{[k]}^n) \\ &\leq \sum_{t=1}^n \left(H(Y_t|X_{S^c t}, U_{[k]}^L) - H(Y_t|U_{[k]}^L, X_{[k]t}) \right) \\ &\leq \sum_{t=1}^n I(X_{St}; Y_t|U_{[k]}^L, X_{S^c t}), \end{aligned}$$

where $X_{St} = (X_{jt})_{j \in S}$. We have

$$p(u_{[k]}^L) = \Pr \{ \forall j \in [k] : W_j \in \mathcal{A}_j(u_{[k]}^L) \} = \prod_{j=1}^k \frac{|\mathcal{A}_j(u_{[k]}^L)|}{|\mathcal{W}_j|}$$

and

$$p(u_{[k]}^L|w_j) = \mathbf{1}\{w_j \in \mathcal{A}_j(u_{[k]}^L)\} \prod_{i \neq j} \frac{|\mathcal{A}_i(u_{[k]}^L)|}{|\mathcal{W}_i|}.$$

Thus

$$p(w_j|u_{[k]}^L) = \frac{p(w_j)p(u_{[k]}^L|w_j)}{p(u_{[k]}^L)} = \frac{\mathbf{1}\{w_j \in \mathcal{A}_j(u_{[k]}^L)\}}{|\mathcal{A}_j(u_{[k]}^L)|}$$

and

$$p(w_{[k]}|u_{[k]}^L) = \frac{p(w_{[k]})p(u_{[k]}^L|w_{[k]})}{p(u_{[k]}^L)} = \frac{\prod_{j=1}^k \mathbf{1}\{w_j \in \mathcal{A}_j(u_{[k]}^L)\}}{\prod_{j=1}^k |\mathcal{A}_j(u_{[k]}^L)|} = \prod_{j=1}^k p(w_j|u_{[k]}^L).$$

Therefore, W_1, \dots, W_k are independent given $U_{[k]}^L$. Recall that at time $t \in [n]$, the output of encoder j is given by $X_{jt} = f_{jt}(W_j, V_j^L)$ for some mapping

$$f_{jt} : [2^{nR_j}] \times \mathcal{V}_j^L \rightarrow \mathcal{X}_j.$$

Also define $U_{0t} = U_{[k]}^L$ for all $t \in [n]$. We have

$$\begin{aligned} p(x_{[k]t}|u_{0t}) &= \sum_{w_{[k]}} p(w_{[k]}|u_{0t})p(x_{[k]t}|w_{[k]}, u_{0t}) \\ &= \sum_{w_{[k]}} \prod_{j=1}^k p(w_j|u_{0t})p(x_{jt}|w_j, u_{0t}) \\ &= \prod_{j=1}^k \sum_{w_j} p(w_j|u_{0t})p(x_{jt}|w_j, u_{0t}) = \prod_{j=1}^k p(x_{jt}|u_{0t}). \end{aligned}$$

Defining a time sharing random variable and applying the usual time sharing argument [13, p. 600] completes the proof.

E. Proposition 7 (The Gaussian MAC)

Consider any $\alpha \in [0, 1/2]$. In the region given in Section V, set $C_{10} = C_{20} = 0$, $C_{1d} = C_{2d} = C_{\text{out}}$, $\rho_1 = \rho_2 = 1$, and

$$\rho_0 = \sqrt{1 - 2^{-4C_{\text{out}}}}.$$

Then the rate pair (R_1^*, R_2^*) given by

$$\begin{aligned} R_1^* &= \frac{1}{2} \log \left(\frac{1 + \gamma_1 + \gamma_2 + 2\rho_0\tilde{\gamma}}{1 + (1 - \rho_0^2)\gamma_2} \right) - C_{\text{out}} \\ R_2^* &= \frac{1}{2} \log (1 + (1 - \rho_0^2)\gamma_2), \end{aligned}$$

is achievable. Since

$$\begin{aligned} C_\alpha(0) &= \alpha \times \frac{1}{2} \log \left(\frac{1 + \gamma_1 + \gamma_2}{1 + \gamma_2} \right) + (1 - \alpha) \times \frac{1}{2} \log(1 + \gamma_2) \\ &= \frac{\alpha}{2} \log(1 + \gamma_1 + \gamma_2) + \frac{1 - 2\alpha}{2} \log(1 + \gamma_2), \end{aligned}$$

we have

$$\begin{aligned} C_\alpha(C_{\text{out}}) - C_\alpha(0) &\geq \alpha R_1^* + (1 - \alpha) R_2^* - C_\alpha(0) \end{aligned} \tag{42}$$

$$= \frac{\alpha}{2} \log \left(1 + \frac{2\rho_0\tilde{\gamma}}{1 + \gamma_1 + \gamma_2} \right) + \frac{1 - 2\alpha}{2} \log \left(1 - \frac{\rho_0^2\gamma_2}{1 + \gamma_2} \right) - C_{\text{out}}. \tag{43}$$

Using the fact that $2^x = 1 + \frac{x}{\log e} + o(x)$ and $\sqrt{1 + o(1)} = 1 + o(1)$, we get

$$\begin{aligned}\rho_0 &= \sqrt{1 - 2^{-4C_{\text{out}}}} \\ &= \sqrt{\frac{4C_{\text{out}}}{\log e} + o(C_{\text{out}})} \\ &= \frac{2}{\sqrt{\log e}} \cdot \sqrt{C_{\text{out}}} + o(\sqrt{C_{\text{out}}}).\end{aligned}$$

In addition,

$$\rho_0^2 = \frac{4C_{\text{out}}}{\log e} + o(C_{\text{out}}) = o(\sqrt{C_{\text{out}}}).$$

Applying $\log(1 + x) = x \log e + o(x)$ to (42) completes the proof for $\alpha \in [0, 1/2]$. The proof for $\alpha \in (1/2, 1]$ follows similarly.

F. Proposition 8 (Capacity region under the CF and conferencing models)

An L -round $(C_{ij})_{i,j=1}^k$ -conference for a blocklength- n code is uniquely determined by a collection of sets $\{\mathcal{W}_{ij}^{(\ell)}\}_{i,j,\ell}$ and mappings

$$h_{ji}^{(\ell)} : [2^{nR_j}] \times \prod_{i': i' \neq j} \mathcal{W}_{i'j}^{\ell-1} \rightarrow \mathcal{W}_{ji}^{(\ell)}$$

where $i, j \in [k]$ and $\ell \in [L]$, and for every $\ell \in [L]$,

$$\mathcal{W}_{ij}^\ell = \prod_{\ell'=1}^{\ell} \mathcal{W}_{ij}^{(\ell')}.$$

Furthermore, the sets $\mathcal{W}_{ij}^{(\ell)}$ satisfy

$$\sum_{\ell \in [L]} \log |\mathcal{W}_{ij}^{(\ell)}| \leq nC_{ij}$$

for all distinct $i, j \in [k]$. Finally, for every message vector (m_1, \dots, m_k) , where $m_j \in [2^{nR_j}]$, define $w_{ji}^{(\ell)}$ recursively as

$$w_{ji}^{(\ell)} = h_{ji}^{(\ell)} \left(m_j, (w_{i'j}^{\ell-1})_{i' \neq j} \right).$$

Our aim is to construct a blocklength- n code for the same MAC with a $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$ -CF that through L rounds of communication with the encoders, provides them with the same information as the L -round conference given above. To this end, for every $j \in [k]$ and $\ell \in [L]$ define the sets $\mathcal{U}_{j\ell}$ and $\mathcal{V}_{j\ell}$ as

$$\begin{aligned}\mathcal{U}_{j\ell} &= \prod_{i: i \neq j} \mathcal{W}_{ji}^{(\ell)} \\ \mathcal{V}_{j\ell} &= \prod_{i: i \neq j} \mathcal{W}_{ij}^{(\ell)}.\end{aligned}$$

Then

$$\begin{aligned}
\sum_{\ell=1}^L \log |\mathcal{U}_{j\ell}| &= \sum_{\ell=1}^L \sum_{i:i \neq j} \log |\mathcal{W}_{ji}^{(\ell)}| \\
&= \sum_{i:i \neq j} \sum_{\ell=1}^L \log |\mathcal{W}_{ji}^{(\ell)}| \\
&\leq n \sum_{i:i \neq j} C_{ji} \leq nC_{\text{in}}^j.
\end{aligned}$$

Similarly, we show

$$\sum_{\ell=1}^L \log |\mathcal{V}_{j\ell}| \leq n \sum_{i:i \neq j} C_{ij} \leq nC_{\text{out}}^j.$$

Next for every $j \in [k]$ and $\ell \in [L]$, define the mapping

$$\begin{aligned}
\varphi_{j\ell} : [2^{nR_j}] \times \mathcal{V}_j^{\ell-1} &\rightarrow \mathcal{U}_{j\ell} \\
(m_j, (w_{ij}^{\ell-1})_{i:i \neq j}) &\mapsto (w_{ji}^{(\ell)})_{i:i \neq j}.
\end{aligned}$$

Similarly, define

$$\begin{aligned}
\psi_{j\ell} : \prod_{i \in [k]} \mathcal{U}_i^\ell &\rightarrow \mathcal{V}_{j\ell} \\
(w_{ij'}^\ell)_{i,j'} &\mapsto (w_{ij}^{(\ell)})_{i:i \neq j}.
\end{aligned}$$

This completes the proof of the first part.

For the second part, we show that the capacity region of a MAC with a single-round $(C_{ij})_{i,j}$ -conference contains the outer bound given in Proposition 5 if $C_{ij} \geq C_{\text{in}}^i$ for all $i, j \in [k]$. The coding strategy is simple. For each $i \in [k]$, encoder i sends the first nC_{in}^i bits of its message to all other encoders. The encoders then form a “common message,” that contains the initial nC_{in}^i bits of message i for all $i \in [k]$. The rest of the proof follows from the forwarding inner bound (Corollary 2) with $C_{i0} = C_{\text{in}}^i$ for all $i \in [k]$.

VIII. CONCLUSION

Cooperative strategies allow for a more efficient allocation of network resources. Here we introduce a model where the encoders of a k -user MAC cooperate through a larger network. This model allows us to construct examples of memoryless networks where removing an edge results in a capacity loss much larger than the capacity of the removed edge, thus proving that the edge removal property [4], [5] does not hold for memoryless networks in general. Finally, we remark that the benefit of cooperation is not limited to achieving higher transmission rates, and cooperative strategies also make networks more reliable. We study the reliability benefit of cooperation in [24].

APPENDIX A

THE MULTIVARIATE COVERING LEMMA

For every positive integer n , define the set $[n] = \{1, \dots, n\}$. Now let k be a positive integer and fix sets $\mathcal{U}_0, \mathcal{U}_1, \dots, \mathcal{U}_{k+1}$. For every nonempty $S \subseteq [k]$ define

$$\mathcal{U}_S = \prod_{j \in S} \mathcal{U}_j.$$

An element of \mathcal{U}_S is denoted with $u_S = (u_j)_{j \in S}$. Let $p(u_0, u_{[k+1]})$ be a probability distribution on the set $\mathcal{U}_0 \times \mathcal{U}_{[k+1]}$. For every $j \in [k]$, let M_j be a nonnegative integer. For every nonempty $S \subseteq [k]$, define the set \mathcal{M}_S as

$$\mathcal{M}_S = \prod_{j \in S} [M_j].$$

and let $\mathcal{M} = \mathcal{M}_{[k]}$. For every $\mathbf{m} = (m_1, \dots, m_k) \in \mathcal{M}$, let the random vector

$$(U_0, U_1(m_1), \dots, U_k(m_k), U_{k+1})$$

have distribution

$$p(u_0) \prod_{j=1}^{k+1} p(u_j | u_0),$$

where $p(u_0)$ and each $p(u_j | u_0)$ are the conditional marginals of $p(u_0, u_{[k+1]})$. In addition, let \mathcal{F} be an arbitrary subset of $\mathcal{U}_0 \times \mathcal{U}_{[k+1]}$. We want to find upper and lower bounds on the probability

$$\Pr \left\{ \forall \mathbf{m} \in \mathcal{M} : (U_0, U_1(m_1), \dots, U_k(m_k), U_{k+1}) \notin \mathcal{F} \right\}.$$

We derive the lower bound (Subsection A-A) using the union bound, which does not depend on the statistical dependencies of the vectors

$$(U_0, U_1(m_1), \dots, U_k(m_k), U_{k+1})$$

for different values of \mathbf{m} . For the upper bound (Subsection A-B), which leads to the multivariate covering lemma, we require a stronger assumption, which we next describe.

Let \mathbf{m} and \mathbf{m}' be in \mathcal{M} . Define the set $S_{\mathbf{m}, \mathbf{m}'}$ as

$$S_{\mathbf{m}, \mathbf{m}'} = \{j \in [k] : m_j = m'_j\}.$$

When \mathbf{m} and \mathbf{m}' are clear from context, we denote $S_{\mathbf{m}, \mathbf{m}'}$ with S . In the proof of the upper bound we require

$$\begin{aligned} & \Pr \left\{ \forall j \in [k] : U_j(m_j) = u_j \text{ and } U_j(m'_j) = u'_j \mid U_0 = u_0, U_{k+1} = u_{k+1} \right\} \\ &= \prod_{j=1}^k p(u_j | u_0) \times \prod_{j \in S^c} p(u'_j | u_0), \end{aligned}$$

for all u_0 and all $(u_j)_j$ and $(u'_j)_j$ such that if $j \in S$, then $u_j = u'_j$ (Assumption I). Note that if there exists a $j \in S$ where $u_j \neq u'_j$ then the probability on the left hand side equals zero.

In the corresponding asymptotic problem (Subsection A-C), we apply our bounds to

$$\Pr \left\{ \forall \mathbf{m} : (U_0^n, U_1^n(m_1), \dots, U_k^n(m_k), U_{k+1}^n) \notin A_\delta^{(n)} \right\},$$

where for every \mathbf{m} ,

$$(U_0^n, U_1^n(m_1), \dots, U_k^n(m_k), U_{k+1}^n)$$

is simply n i.i.d. copies of the original random vector

$$(U_0, U_1(m_1), \dots, U_k(m_k), U_{k+1}),$$

(Assumption II) and $A_\delta^{(n)}$ is the weakly typical set [13, p. 521] defined with respect to the distribution $p(u_0, u_{[k+1]})$.

The multivariate covering lemma follows.

Lemma 16 (Multivariate Covering Lemma). *Suppose assumptions (I) and (II) hold for the joint distribution of*

$$U_0^n, \{U_1^n(m_1), \dots, U_k^n(m_k)\}_{\mathbf{m}}, U_{k+1}^n.$$

For the direct part, suppose for all $j \in [k]$, $M_j \geq 2^{nR_j}$. If for all nonempty $S \subseteq [k]$,

$$\sum_{j \in S} R_j > \sum_{j \in S} H(U_j|U_0) - H(U_S|U_0, U_{k+1}) + (8k - 2|S| + 10)\delta, \quad (44)$$

then

$$\lim_{n \rightarrow \infty} \Pr \left\{ \exists \mathbf{m} : (U_0^n, U_1^n(m_1), \dots, U_k^n(m_k), U_{k+1}^n) \in A_\delta^{(n)} \right\} = 1. \quad (45)$$

For the converse, assume for all $j \in [k]$, $M_j \leq 2^{nR_j}$. If (45) holds, then

$$\sum_{j \in S} R_j \geq \sum_{j \in S} H(U_j|U_0) - H(U_S|U_0, U_{k+1}) - 2(|S| + 1)\delta,$$

for all nonempty $S \subseteq [k]$.

Remark. In the direct part of Lemma 16, we can weaken the lower bound on $\sum_{j \in S} R_j$ when $S = [k]$. Specifically, we can replace (44) with

$$\sum_{j=1}^k R_j > \sum_{j=1}^k H(U_j|U_0) - H(U_{[k]}|U_0, U_{k+1}) + 2(k+1)\delta.$$

for $S = [k]$.

A. The Lower Bound

Define the distribution $p_{\text{ind}}(u_0, u_{[k+1]})$ on the set $\mathcal{U}_0 \times \mathcal{U}_{[k+1]}$ as

$$p_{\text{ind}}(u_0, u_{[k+1]}) = p(u_0, u_{k+1}) \prod_{j \in [k]} p(u_j|u_0).$$

For every $S \subseteq [k]$, define \mathcal{F}_S as the projection of \mathcal{F} on $\mathcal{U}_0 \times \mathcal{U}_S \times \mathcal{U}_{k+1}$, and for every $(u_0, u_S, u_{k+1}) \in \mathcal{F}_S$, let $\mathcal{F}(u_0, u_S, u_{k+1})$ be the set of all u_{S^c} such that $(u_0, u_{[k+1]}) \in \mathcal{F}$. In addition, for every nonempty $S \subseteq [k]$, let α_S and β_S be constants such that for all $(u_0, u_S, u_{k+1}) \in \mathcal{F}_S$

$$\alpha_S \leq \log \frac{p(u_S|u_0, u_{k+1})}{p_{\text{ind}}(u_S|u_0)},$$

and for all $(u_0, u_S, u_{S^c}, u_{k+1}) \in \mathcal{F}$,

$$\beta_S \leq \log \frac{p(u_S|u_0, u_{S^c}, u_{k+1})}{p_{\text{ind}}(u_S|u_0)}.$$

Furthermore, let the constant γ satisfy

$$\gamma \geq \log \frac{p(u_{[k]}|u_0, u_{k+1})}{p_{\text{ind}}(u_{[k]}|u_0)}$$

for all $(u_0, u_{[k]}, u_{k+1}) \in \mathcal{F}$.

For every $\mathbf{m} = (m_1, \dots, m_k) \in \mathcal{M}$, define the random variable $Z_{\mathbf{m}}$ as

$$Z_{\mathbf{m}} = \mathbf{1} \left\{ (U_0, U_1(m_1), \dots, U_k(m_k), U_{k+1}) \in \mathcal{F} \right\}$$

and set

$$Z = \sum_{\mathbf{m} \in \mathcal{M}} Z_{\mathbf{m}}.$$

Our aim is to find a lower bound for $\Pr\{Z = 0\}$. Note that for every nonempty $S \subseteq [k]$,

$$\begin{aligned} \Pr\{\exists \mathbf{m} : Z_{\mathbf{m}} = 1\} &= \Pr\left\{\exists \mathbf{m} : (U_0, U_1(m_1), \dots, U_k(m_k), U_{k+1}) \in \mathcal{F}\right\} \\ &\leq \Pr\left\{\exists \mathbf{m} : (U_0, (U_j(m_j))_{j \in S}, U_{k+1}) \in \mathcal{F}_S\right\} \\ &\leq |\mathcal{M}_S| \sum_{\mathcal{F}_S} p(u_0, u_{k+1}) p_{\text{ind}}(u_S | u_0) \\ &\leq |\mathcal{M}_S| 2^{-\alpha_S} \sum_{\mathcal{F}_S} p(u_0, u_S, u_{k+1}) \\ &\leq |\mathcal{M}_S| 2^{-\alpha_S}. \end{aligned}$$

Thus

$$\begin{aligned} \Pr\{Z = 0\} &= 1 - \Pr\{\exists \mathbf{m} : Z_{\mathbf{m}} = 1\} \\ &\geq 1 - \min_{|S| \neq \emptyset} |\mathcal{M}_S| 2^{-\alpha_S}. \end{aligned} \tag{46}$$

B. The Upper Bound

In deriving our upper bound on $\Pr\{Z = 0\}$, we apply conditioning and Chebyshev's inequality. Thus, the factor

$$\frac{1}{\left(\Pr\{\mathcal{F}(u_0, u_{k+1})\}\right)^2}$$

appears, where

$$\begin{aligned} \Pr\{\mathcal{F}(u_0, u_{k+1})\} &= \Pr\left\{U_{[k]} \in \mathcal{F}(u_0, u_{k+1}) \mid U_0 = u_0, U_{k+1} = u_{k+1}\right\} \\ &= \sum_{u_{[k]} \in \mathcal{F}(u_0, u_{k+1})} p(u_{[k]} | u_0, u_{k+1}) \end{aligned}$$

and $\mathcal{F}(u_0, u_{k+1})$ (Subsection A-A) is simply the set of all $u_{[k]} \in \mathcal{U}_{[k]}$ that satisfy $(u_0, u_{[k]}, u_{k+1}) \in \mathcal{F}$. Thus to get a reasonably accurate upper bound, we require $\Pr\{\mathcal{F}(u_0, u_{k+1})\}$ to be large. However, as we cannot guarantee this for all (u_0, u_{k+1}) , we partition the (u_0, u_{k+1}) pairs into “good” and “bad” sets, corresponding to large and small values of $\Pr\{\mathcal{F}(u_0, u_{k+1})\}$, respectively. The probability of the good set is large when $\Pr\{(U_0, U_{[k+1]}) \in \mathcal{F}\}$ is sufficiently large. To see this, fix $\epsilon > 0$. Following Appendix III of [12], define the set $\mathcal{G} \subseteq \mathcal{U}_0 \times \mathcal{U}_{k+1}$ as

$$\mathcal{G} = \{(u_0, u_{k+1}) : \Pr\{\mathcal{F}(u_0, u_{k+1})\} \geq 1 - \epsilon\},$$

Note that \mathcal{G} is the set of all good (u_0, u_{k+1}) pairs as defined above. We have

$$\begin{aligned} \Pr\{(U_0, U_{[k+1]}) \in \mathcal{F}\} &= \sum_{u_0, u_{k+1}} p(u_0, u_{k+1}) \Pr\{\mathcal{F}(u_0, u_{k+1})\} \\ &\leq (1 - \epsilon) \Pr\{(U_0, U_{k+1}) \notin \mathcal{G}\} + \Pr\{(U_0, U_{k+1}) \in \mathcal{G}\} \\ &= 1 - \epsilon \Pr\{(U_0, U_{k+1}) \notin \mathcal{G}\}. \end{aligned}$$

Thus

$$\Pr\{(U_0, U_{k+1}) \notin \mathcal{G}\} \leq \frac{1}{\epsilon} \Pr\{(U_0, U_{[k+1]}) \notin \mathcal{F}\}. \quad (47)$$

Our aim is to find an upper bound for $\Pr\{Z = 0\}$. To do this, we write

$$\begin{aligned} \Pr\{Z = 0\} &= \sum_{u_0, u_{k+1}} p(u_0, u_{k+1}) \Pr\{Z = 0 | u_0, u_{k+1}\} \\ &\leq \frac{1}{\epsilon} \Pr\{(U_0, U_{[k]}, U_{k+1}) \notin \mathcal{F}\} + \sum_{(u_0, u_{k+1}) \in \mathcal{G}} p(u_0, u_{k+1}) \Pr\{Z = 0 | u_0, u_{k+1}\}, \end{aligned} \quad (48)$$

where the inequality follows from (47). Therefore, to find an upper bound on $\Pr\{Z = 0\}$, it suffices to find an upper bound on $\Pr\{Z = 0 | u_0, u_{k+1}\}$ for all $(u_0, u_{k+1}) \in \mathcal{G}$.

Fix $(u_0, u_{k+1}) \in \mathcal{G}$. We use Chebyshev's inequality to find an upper bound on $\Pr\{Z = 0 | u_0, u_{k+1}\}$. Thus we need to calculate $\mathbb{E}[Z | u_0, u_{k+1}]$ and $\mathbb{E}[Z^2 | u_0, u_{k+1}]$. For a given \mathbf{m} , from the definition of γ (Subsection A-A) it follows

$$\begin{aligned} \mathbb{E}[Z_{\mathbf{m}} | u_0, u_{k+1}] &= \Pr\{(U_1(m_1), \dots, U_k(m_k)) \in \mathcal{F}(u_0, u_{k+1}) | u_0, u_{k+1}\} \\ &= \sum_{\mathcal{F}(u_0, u_{k+1})} p_{\text{ind}}(u_{[k]} | u_0) \\ &\geq \sum_{\mathcal{F}(u_0, u_{k+1})} 2^{-\gamma} p(u_{[k]} | u_0, u_{k+1}) \\ &= 2^{-\gamma} \Pr\{\mathcal{F}(u_0, u_{k+1})\} \geq (1 - \epsilon) 2^{-\gamma}. \end{aligned}$$

where the last inequality follows from the fact that $(u_0, u_{k+1}) \in \mathcal{G}$. Thus, by linearity of expectation,

$$\mathbb{E}[Z | u_0, u_{k+1}] \geq |\mathcal{M}| 2^{-\gamma} (1 - \epsilon). \quad (49)$$

Next, we find an upper bound on $\mathbb{E}[Z^2 | u_0, u_{k+1}]$. We have

$$Z^2 = \sum_{\mathbf{m}} Z_{\mathbf{m}}^2 + \sum_{\mathbf{m} \neq \mathbf{m}'} Z_{\mathbf{m}} Z_{\mathbf{m}'} = Z + \sum_{\mathbf{m} \neq \mathbf{m}'} Z_{\mathbf{m}} Z_{\mathbf{m}'},$$

since $Z_{\mathbf{m}}^2 = Z_{\mathbf{m}}$ and $Z = \sum_{\mathbf{m}} Z_{\mathbf{m}}$. Thus

$$\mathbb{E}[Z^2 | u_0, u_{k+1}] = \mathbb{E}[Z | u_0, u_{k+1}] + \mathbb{E}\left[\sum_{\mathbf{m} \neq \mathbf{m}'} Z_{\mathbf{m}} Z_{\mathbf{m}'} \middle| u_0, u_{k+1}\right]$$

For any pair of distinct \mathbf{m} and \mathbf{m}' with nonempty $S = S_{\mathbf{m}, \mathbf{m}'}$, we have

$$\begin{aligned} &\mathbb{E}[Z_{\mathbf{m}} Z_{\mathbf{m}'} | u_0, u_{k+1}] \\ &= \sum_{\mathcal{F}_S(u_0, u_{k+1})} p_{\text{ind}}(u_S | u_0) \left[\sum_{u_{S^c} \in \mathcal{F}(u_0, u_S, u_{k+1})} p_{\text{ind}}(u_{S^c} | u_0) \right]^2 \\ &\leq 2^{-\alpha_S - 2\beta_{S^c}} \sum_{\mathcal{F}_S(u_0, u_{k+1})} p(u_S | u_0, u_{k+1}) \left[\sum_{u_{S^c} \in \mathcal{F}(u_0, u_S, u_{k+1})} p(u_{S^c} | u_0, u_S, u_{k+1}) \right]^2 \\ &\leq 2^{-\alpha_S - 2\beta_{S^c}}, \end{aligned}$$

where $\mathcal{F}_S(u_0, u_{k+1})$ is the set of all u_S that satisfy $(u_0, u_S, u_{k+1}) \in \mathcal{F}_S$. On the other hand, if $S = S_{\mathbf{m}, \mathbf{m}'}$ is empty, then $Z_{\mathbf{m}}$ and $Z'_{\mathbf{m}'}$ are independent given $(U_0, U_{k+1}) = (u_0, u_{k+1})$, and

$$\mathbb{E}[Z_{\mathbf{m}} Z'_{\mathbf{m}'} | u_0, u_{k+1}] = (\mathbb{E}[Z_{\mathbf{m}} | u_0, u_{k+1}])^2.$$

Thus

$$\begin{aligned} \mathbb{E}[Z^2 | u_0, u_{k+1}] &= \mathbb{E}[Z | u_0, u_{k+1}] + (\mathbb{E}[Z | u_0, u_{k+1}])^2 \\ &\quad + \sum_{\emptyset \subset S \subset [k]} |\mathcal{M}_S| \prod_{j \in S^c} (|\mathcal{M}_j|^2 - |\mathcal{M}_j|) \mathbb{E}[Z_{\mathbf{m}} Z'_{\mathbf{m}'} | u_0, u_{k+1}] \\ &\leq \mathbb{E}[Z | u_0, u_{k+1}] + (\mathbb{E}[Z | u_0, u_{k+1}])^2 + \sum_{\emptyset \subset S \subset [k]} |\mathcal{M}_S| |\mathcal{M}_{S^c}|^2 2^{-\alpha_S - 2\beta_{S^c}}, \end{aligned} \quad (50)$$

where the notation $\emptyset \subset S \subset [k]$ means that S is a nonempty proper subset of $[k]$. Thus for all $(u_0, u_{k+1}) \in \mathcal{G}$, we have

$$\begin{aligned} \Pr\{Z = 0 | u_0, u_{k+1}\} &\leq \Pr\left\{|Z - \mathbb{E}[Z | u_0, u_{k+1}]| \geq \mathbb{E}[Z | u_0, u_{k+1}] \mid u_0, u_{k+1}\right\} \\ &\stackrel{(a)}{\leq} \frac{\text{Var}(Z | u_0, u_{k+1})}{(\mathbb{E}[Z | u_0, u_{k+1}])^2} = \frac{\mathbb{E}[Z^2 | u_0, u_{k+1}]}{(\mathbb{E}[Z | u_0, u_{k+1}])^2} - 1 \\ &\stackrel{(b)}{\leq} \frac{1}{1 - \epsilon} |\mathcal{M}|^{-1} 2^\gamma + \frac{1}{(1 - \epsilon)^2} \sum_{\emptyset \subset S \subset [k]} |\mathcal{M}_S|^{-1} 2^{-\alpha_S - 2\beta_{S^c} + 2\gamma}, \end{aligned}$$

where (a) follows from Chebyshev's inequality and (b) follows from (49) and (50). Now using (48), we get

$$\Pr\{Z = 0\} \leq \frac{1}{\epsilon} \Pr\{\mathcal{F}^c\} + \frac{1}{1 - \epsilon} |\mathcal{M}|^{-1} 2^\gamma + \frac{1}{(1 - \epsilon)^2} \sum_{\emptyset \subset S \subset [k]} |\mathcal{M}_S|^{-1} 2^{-\alpha_S - 2\beta_{S^c} + 2\gamma}. \quad (51)$$

C. The Asymptotic Result

In this section, using our lower and upper bounds, we prove Lemma 16. We first prove the direct part using our upper bound from Section A-B. Set $\mathcal{F} = A_\delta^{(n)}$ and for every $j \in [k]$, choose an integer $M_j \geq 2^{nR_j}$. Choose a sequence $\{\epsilon_n\}_n$ such that

$$\lim_{n \rightarrow \infty} \frac{1}{\epsilon_n} \Pr\{(A_\delta^{(n)})^c\} = 0.$$

Fix a nonempty $S \subseteq [k]$. Notice that if $(U_0^n, (U_j^n)_{j \in S}, U_{k+1}^n) \in \mathcal{F}_S$, then

$$\left| \log \frac{p(u_S^n | u_0^n, u_{k+1}^n)}{\prod_{j \in S} p(u_j^n | u_0^n)} - n \left(\sum_{j \in S} H(U_j | U_0) - H(U_S | U_0, U_{k+1}) \right) \right| \leq 2n(|S| + 1)\delta.$$

Thus we may choose

$$\alpha_S = n \left(\sum_{j \in S} H(U_j | U_0) - H(U_S | U_0, U_{k+1}) - 2(|S| + 1)\delta \right)$$

and

$$\gamma = n \left(\sum_{j=1}^k H(U_j | U_0) - H(U_{[k]} | U_0, U_{k+1}) + 2(k + 1)\delta \right).$$

Similarly, for every nonempty $S \subseteq [k]$, we choose β_S as

$$\beta_S = n \left(\sum_{j \in S} H(U_j | U_0) - H(U_S | U_0, U_{S^c}, U_{k+1}) - 2(|S| + 1)\delta \right),$$

since for every $(U_0^n, (U_j^n)_{j \in S}, (U_j^n)_{j \in S^c}) \in \mathcal{F}$,

$$\left| \log \frac{p(u_S^n | u_0^n, u_{S^c}^n, u_{k+1}^n)}{\prod_{j \in S} p(u_j^n | u_0^n)} - n \left(\sum_{j \in S} H(U_j | U_0) - H(U_S | U_0, U_{S^c}, U_{k+1}) \right) \right| \leq 2n(|S| + 1)\delta.$$

From our upper bound, Equation (51), it now follows that if for all nonempty $S \subset [k]$,

$$\begin{aligned} \sum_{j \in S} R_j &> \frac{1}{n}(2\gamma - \alpha_S - 2\beta_{S^c}) \\ &= 2 \sum_{j=1}^k H(U_j | U_0) - 2H(U_{[k]} | U_0, U_{k+1}) - \sum_{j \in S} H(U_j | U_0) + H(U_S | U_0, U_{k+1}) \\ &\quad - 2 \sum_{j \in S^c} H(U_j | U_0) + 2H(U_{S^c} | U_0, U_S, U_{k+1}) + (8k - 2|S| + 10)\delta \\ &= \sum_{j \in S} H(U_j | U_0) - H(U_S | U_0, U_{k+1}) + (8k - 2|S| + 10)\delta, \end{aligned}$$

and for $S = [k]$,

$$\sum_{j=1}^k R_j > \frac{1}{n}\gamma = \sum_{j=1}^k H(U_j | U_0) - H(U_{[k]} | U_0, U_{k+1}) - 2(k+1)\delta,$$

then

$$\lim_{n \rightarrow \infty} \Pr \left\{ \exists \mathbf{m} : (U_0^n, U_1^n(m_1), \dots, U_k^n(m_k), U_{k+1}^n) \in A_\delta^{(n)} \right\} = 1. \quad (52)$$

Next we prove the converse. Suppose for each $j \in [k]$, $M_j \leq 2^{nR_j}$ and (52) holds. Then from our lower bound, Equation (46), it follows

$$\sum_{j \in S} R_j \geq \frac{1}{n}\alpha_S = \sum_{j \in S} H(U_j | U_0) - H(U_S | U_0, U_{k+1}) - 2(|S| + 1)\delta,$$

for all nonempty $S \subseteq [k]$.

APPENDIX B LARGE DEVIATIONS

In this appendix, we state and prove the following result. It is well known and is included for completeness.

Lemma 17. *Choose a distribution $p(u_{[k]})$ on the alphabet $\mathcal{U}_{[k]}$, which may be continuous or discrete. Suppose there exists $t_0 > 0$ so that for all nonempty $S \subseteq [k]$ and $t \in (-t_0, t_0)$,*

$$\mathbb{E}[p(U_S)^{-t}] < \infty.$$

Then there exists a nondecreasing function $I : \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}$ such that for all sufficiently large n ,

$$\Pr \{ A_\epsilon^{(n)}(U_{[k]}) \} \geq 1 - 2^{-nI(\epsilon)}.$$

Proof: The moment generating function of a random variable X is defined as

$$M(t) = \mathbb{E}[e^{tX}]$$

for all real t for which the expectation on the right hand side is finite. If M is defined on a neighborhood of 0, say $(-t_1, t_1)$ for some $t_1 > 0$, then it has a Taylor series expansion with a positive radius of convergence [25, pp. 278-280]. In particular,

$$\frac{d}{dt}M(t)\big|_{t=0} = \mathbb{E}[X].$$

We next find an upper bound for $\Pr\{X \geq a\}$ for any $a \in \mathbb{R}$. Choose $t \in (0, t_1)$. Using Markov's inequality, we get

$$\begin{aligned} \Pr\{X \geq a\} &= \Pr\{tX \geq ta\} \\ &= \Pr\{e^{tX} \geq e^{ta}\} \\ &\leq e^{-ta} \mathbb{E}[e^{tX}] \\ &= e^{\log M(t) - ta} \end{aligned}$$

Since $t \in (0, t_1)$ was arbitrary, we get

$$\Pr\{X \geq a\} \leq e^{\inf_{t \in (0, t_1)} (\ln M(t) - ta)}.$$

Define the function f as

$$f(t) = \ln M(t) - ta.$$

Then $f(0) = 0$ and $f'(0) = \mathbb{E}[X] - a$. Thus if $a > \mathbb{E}[X]$,

$$\inf_{t \in (0, t_1)} (\ln M(t) - ta) < 0. \quad (53)$$

If we apply the same inequality to the random variable

$$\frac{1}{n} \sum_{i=1}^n X_i,$$

where the X_i 's are i.i.d. copies of X , we get

$$\Pr\left\{\sum_{i=1}^n X_i \geq na\right\} \leq e^{n \inf_{t \in (0, t_1)} (\ln M(t) - ta)}. \quad (54)$$

Now consider a random vector (U_1, \dots, U_k) with distribution $p(u_1, \dots, u_k)$. For every nonempty $S \subseteq [k]$, let U_S denote the random vector $(U_j)_{j \in S}$. Let (U_1^n, \dots, U_k^n) be n i.i.d. copies of (U_1, \dots, U_k) . By applying inequality (54) to the random variables $\{\log \frac{1}{p(U_{Si})}\}_{i=1}^n$ and setting $a = H(U_S) + \epsilon$ for some $\epsilon > 0$, we get

$$\Pr\left\{\sum_{i=1}^n \log \frac{1}{p(U_{Si})} \geq n(H(U_S) + \epsilon)\right\} \leq 2^{-nI_S(\epsilon)}, \quad (55)$$

where $I_S(\epsilon)$ is given by

$$I_S(\epsilon) = \inf_{t \in (0, t_0)} \left\{ \frac{t}{\ln 2} (H(U_S) + \epsilon) - \log \mathbb{E}[p(U_S)^{-t}] \right\} \quad (56)$$

Let

$$I(\epsilon) = \frac{1}{2} \min_{S \subseteq [k]} I_S(\epsilon).$$

By the union bound we get

$$\begin{aligned}
 \Pr \{ (U_1^n, \dots, U_k^n) \notin A_\epsilon^{(n)}(U_1, \dots, U_k) \} &\leq 2 \sum_{\emptyset \subsetneq S \subseteq [k]} e^{-n I_S(\epsilon)} \\
 &\leq 2(2^k - 1) 2^{-n \min_S I_S(\epsilon)} \\
 &\leq 2^{-n I(\epsilon)},
 \end{aligned}$$

where the last inequality holds for all sufficiently large n . Finally, note that since by (53) and (56), each $I_S(\epsilon)$ is positive and nondecreasing, so is $I(\epsilon)$. ■

ACKNOWLEDGMENT

The first author thanks M. F. Wong for useful discussions regarding the proof of Theorem 3.

REFERENCES

- [1] P. Noorzad, M. Effros, and M. Langberg, "On the cost and benefit of cooperation," in *Proc. IEEE Int. Symp. Information Theory*, 2015.
- [2] —, "The unbounded benefit of encoder cooperation for the k -user MAC," in *Proc. IEEE Int. Symp. Information Theory*, 2016.
- [3] M. L. Ulrey, "The capacity region of a channel with s senders and r receivers," *Inform. Contr.*, vol. 29, pp. 185–203, 1975.
- [4] T. Ho, M. Effros, and S. Jalali, "On equivalence between network topologies," in *Proc. Allerton Conf. Communication, Control, and Computing*, 2010.
- [5] S. Jalali, M. Effros, and T. Ho, "On the impact of a single edge on the network coding capacity," in *Information Theory and Applications Workshop*, 2011.
- [6] P. Noorzad, M. Effros, M. Langberg, and T. Ho, "On the power of cooperation: Can a little help a lot?" in *Proc. IEEE Int. Symp. Information Theory*, 2014.
- [7] T. S. Han, "The capacity region of general multiple-access channel with certain correlated sources," *Inform. Contr.*, vol. 40, pp. 37–60, 1979.
- [8] F. M. J. Willems, "The discrete memoryless multiple access channel with partially cooperating encoders," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 3, pp. 441–445, 1983.
- [9] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 3, pp. 306–311, 1979.
- [10] A. El Gamal and E. C. van der Meulen, "A proof of Marton's coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 1, pp. 120–122, 1981.
- [11] A. El Gamal and Y.-H. Kim, *Network Information Theory*, 2nd ed. Cambridge University Press, 2012.
- [12] R. Koetter, M. Effros, and M. Médard, "A theory of network equivalence—Part II: Multiterminal channels," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 3709–3732, 2014.
- [13] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley, 2006.
- [14] R. Ahlswede, "Multi-way communication channels," in *Proc. Int. Symp. Information Theory*, 1971.
- [15] —, "The capacity region of a channel with two senders and two receivers," *Ann. Prob.*, vol. 2, pp. 805–814, 1974.
- [16] H. Liao, "Multiple access channels," Ph.D. dissertation, Dept. Elect. Eng., Univ. Hawaii, 1972.
- [17] O. Simeone, O. Somekh, G. Kramer, H. V. Poor, and S. Shamai, "Three-user Gaussian multiple access channel with partially cooperating encoders," in *Asilomar Conf. on Signals, Systems, and Computers*, 2008.
- [18] T. M. Cover, A. El Gamal, and M. Salehi, "Multiple access channels with arbitrary correlated sources," *IEEE Trans. Inf. Theory*, vol. IT-26, no. 6, pp. 648–657, 1980.
- [19] J. D. Hunter, "Matplotlib: A 2D graphics environment," *IEEE Comput. Sci. Eng.*, vol. 9, no. 3, pp. 90–95, 2007.
- [20] A. Schrijver, *Combinatorial Optimization: Polyhedra and Efficiency*. Springer, 2003, vol. B.
- [21] S. Watanabe, "Information theoretical analysis of multivariate correlation," *IBM J. Res. Dev.*, vol. 4, pp. 66–82, 1960.
- [22] R. F. Bass, *Real Analysis for Graduate Students*, 2nd ed., 2013.

- [23] L. Pardo, *Statistical Inference Based on Divergence Measures*. Chapman & Hall/CRC, 2006.
- [24] P. Noorzad, M. Effros, and M. Langberg, “Can negligible cooperation increase network reliability?” 2016. [Online]. Available: <http://arxiv.org/abs/1601.05769>
- [25] P. Billingsley, *Probability and Measure*, 3rd ed. SIAM, 1995.